

Helping Soviet science (continued)

At last there is some inclination in the West to help ex-Soviet science over the hurdles ahead, but plans now afoot are probably insufficient, and could yet be too late.

SOMETHING seems at last to be happening in the West to help safeguard ex-Soviet science from the hazards ahead, but painfully slowly (see page 182). It is a curious business; there is general agreement that even self-interest predicates action of some kind, but no general agreement on what kind of action that should be.

Much the same is true on the wider question of what should be done to help the Commonwealth of Independent States (CIS) as a whole. It has come to something that it should have taken a speech last week by ex-president Richard M. Nixon of the United States to put that question where it belongs, at the front of the minds of politicians otherwise preoccupied with their own re-election prospects. Yet, if they listen, there is a risk of subsuming the needs of ex-Soviet science in the more general question of how best to help keep most CIS members pointing in the directions they are now following. Although the issues are linked, they are also distinct.

Nixon's argument repeats one often made since Mr Mikhail Gorbachev became general secretary of the Soviet Communist Party at the beginning of 1985: if tendencies towards enlightenment and reform in the East are snuffed out by a return to the bad old ways, it will be the worse for everybody. Much was heard of this in July last year, when Gorbachev attended the meeting of the G-7 industrialized governments in London — and was offered associate membership of the International Monetary Fund, but neither full membership nor a handsome cheque. Nixon is right to say that the present need for help is even more urgent; the Yeltsin government has gone irretrievably far down the road of economic change, but nobody can yet tell whether the change will bring lasting reform, while the break-up of the old Soviet monolith has raised fears that a conflict between, say, Russia and Ukraine would not be as easily confined to where it started as has been that between Armenia and Azerbaijan. That begs the question of what kind of help would help, but large cheques are probably indispensable.

The case for consideration for ex-Soviet science has some components in common with that argument, but others that are not. Its essence is that science is a collective enterprise, to which ex-Soviet and predominantly Russian researchers continue to make important contributions. Can the rest of the scientific enterprise in good conscience let those concerned be squeezed out by the

privations that lie ahead? Especially when the hard-currency cost of temporary rescue is so small? And it is not just that the army of those now at the bench includes an ample sprinkling of able people, but that if the able were to take the easy course and move elsewhere, their successors would be less well educated and might not even exist.

But why do the states of the CIS, and Russia particularly, deserve special help? One of the ironies of the present state of affairs is that the relatively good fortune of ex-Soviet science owes much to the support provided for research by the previous regime, now mostly scorned. But the origins of the ability (if not yet the accomplishment) of ex-Soviet science are irrelevant to an assessment of its needs. And, naturally, none of this implies that researchers elsewhere in Central Europe deserve less sympathetic consideration, but the danger of social collapse is less immediate there. Nor is ex-Soviet science powerless to help itself; symbolically at least, it would help if the Russian Academy of Sciences, the successor of the old Soviet Academy, were to turn itself into the smaller modern academy the states of the CIS need.

It is also natural that there should be many in the West who insist that it would be easier to do something for Soviet science if there were already funds enough to spend at home. "Does not", the doubters ask, "charity begin at home?" (On the broader question, President George Bush last week responded to Nixon by saying that the US Congress had not given him a blank cheque.) The fallacy is that charity is not involved; the benefit is the preservation of an international enterprise whose value is greater than the sum of its parts. And the cost, in present circumstances, would be very small.

What, against this background, should be made of the schemes for helping ex-Soviet science now being put in place? The plan canvassed at last week's meeting of OECD science ministers, which is artfully designed (at a cost of \$50 million) to keep people skilled at weapons design off the international military market, should also help to bring an important part of the ex-Soviet enterprise into the general swim of things, but nobody should be surprised if its public purpose proves elusive; people inclined to help others make, say, nuclear weapons are unlikely to be kept at home by the modest rewards now offered.

The willingness of grant-making agencies in the West



to allow grant recipients to spend up to 5 per cent of their awards on collaboration with people in Europe and further East could be even more valuable, for such funds are most likely to be spent where collaboration is most likely to flourish. Schemes for keeping ex-Soviet research groups in being with the help of research contracts with commercial organizations in the West — the US Department of Energy and Sun Microsystems are two examples — could be just as valuable, especially if they help to engender more of the commercial sense that Central Europe needs. Sadly, more-academic organizations have been much slower off the mark. Several learned academies have issued statements drawing attention to the need, but have mostly confined themselves to a re-examination of existing exchange schemes. And the chief tangible result of several months of brooding by the larger private foundations (mostly in the United States) has been the decision of the MacArthur Foundation to spend \$3 million a year and to open an office in Moscow. Is that enough? □

Living with monopoly

Governments everywhere are selling nationalized industries to private persons, but problems of monopoly persist.

THE governments of Central Europe now bent on selling nationalized industries to private investors, but not quite sure how that should be done, might usefully read a series of public documents put out in Britain in the past few weeks. A decade ago, the British government was a pioneer in the field, and may even have been responsible for the word 'privatization'. Now, with the passage of time, it is inevitable that people should be looking back to see how well the job was done. Hindsight gives a somewhat jaundiced view.

So much is clear from what the House of Commons Select Committee on Energy had to say last week about one of the more recent disposals, the sale of the nationalized electricity industry in 1989. The essence of the difficulty is that the electricity industry used to be a strict monopoly — three organizations (two in Scotland) had an exclusive right to generate and sell electricity in their exclusive territories. Part of the case for selling them was the fear that their monopoly rights had made them inefficient. (Another was the need for cash, used to pay off public debt.) But how to create an electricity industry that does not enjoy some kind of monopoly, at least in the sale of kilowatt-hours to those who use them? There is no obvious way.

The British government's solution was to split what was a vertically integrated industry into three horizontal layers — generators of electricity, distributors (who operate the National Grid) and retailers (who deal with the consumers of electricity, but who have the right to generate as well if they choose).

One of the select committee's worries is that the system which has emerged embodies almost no effective compe-

tion, but that is hardly a surprise: the retail organizations remain monopolies in their territories, and are regulated by restraints on the prices they are allowed to charge. Moreover, there is a now a tendency for these local monopolies to become more self-contained; there is a fashion for building relatively cheap gas-burning generating sets that will insulate them from the high cost of electricity drawn from the National Grid at times of peak demand. But what else does commercial logic dictate? The committee's worry that the three generating companies (one large, one smaller and one that operates Britain's nuclear capacity) are able to manipulate the prices paid by the National Grid for electricity is more substantial, but may disappear with the passage of time; in principle, at least, anybody believing that electricity can be generated cheaply can join the fray.

This is what seems to be happening with telecommunications, the first substantial privatization in Britain. Although hopes have been disappointed that British Telecom, with more than 90 per cent of Britain's business, would be put under pressure by competition from the company called Mercury, what seems most likely to untie the monopoly knot is last year's decision that the British Telecom network can be used by any provider who wishes, and who is prepared to pay a regulated fee for the privilege. Enough providers of, say, long-distance calls would soon create the pressure on British Telecom that the public interest needs. (Piquantly, still-nationalized British Rail is apparently anxious to enter the lists.)

Meanwhile, there is much that could be done by regulatory fiat. More than a month ago, the industry's regulator, Sir Bryan Carsberg, put out a document reiterating the decision that British Telecom (and others) should be regulated by the prices they actually charge, and not by some other yardstick such as the rate of return on capital. Over the short run, there is force in the argument that price control is the most economic way of keeping monopolies on their toes, and restraining innate rapacity. (Over the longer run, private monopolies have the freedom of knowing that even the toughest regulators are unlikely ever to drive them into bankruptcy.) But Carsberg says he is looking for mechanisms of charging for telecommunications services that will be inherently efficient. Why not simply give all users of, say, telephones, a choice of tariffs allowing them to choose for themselves the proportions of their bills spent on fixed charges and those that are proportional to usage? That way, users would benefit from charges matched more accurately to their needs while the monopoly's capital investment would be more fully used.

In Central Europe, such matters may seem refinements, but they are issues inseparable from the management of private monopolies. Sooner or later, they are inescapable, as the British government has now discovered. And as the United States and Germany (with electricity production) have discovered, it is irrelevant whether the industries concerned begin as nationalized concerns. □

Japanese researchers rule out gene patents

- Stance opposes US, UK positions
- Plan to make data freely available

Osaka

THE leading scientists in Japan's effort to sequence the genomes of humans and plants say they do not plan to join the United States and Britain in filing patents on their discoveries but will instead publish the data and make it freely available. The various government ministries and agencies involved in genome research have yet to develop a policy on the patenting of genes, and it would be unprecedented for the government to intervene in a scientific decision of this type.

A controversy erupted among genome researchers last year after Craig Venter of the US National Institutes of Health (NIH) filed patent applications for more than 300 cDNA fragments of human genes. Researchers at Britain's Medical Research Council (MRC) initially condemned Venter's move, but the MRC has since decided to file patent applications for more than 1,000 cDNA sequences.

At present, only the Science and Technology Agency, which is expected to play a leading role in Japan's human genome project, has established a small committee of five lawyers to examine the issue of patenting human genes. But the committee is not expected to produce a report for at least a year.

Kenichi Matsubara, who works at the Institute of Molecular and Cellular Biology of Osaka University, says that he will not file patents for the more than 2,000 sequences of human complementary DNA (cDNA) that his group currently has deposited in the DNA databank of Japan in Mishima. Matsubara says he will make the sequences available once he has published the work, which he expects to do soon.

Matsubara says there is no possibility of his decision being reversed by the Japa-

nese government. The Ministry of Education, Science and Culture, which funds the Osaka group, leaves it up to individual scientists and their universities to decide whether to file patents.

The Japanese group uses sequencing techniques that are similar to what is done by the NIH team, but the goals of the two teams differ somewhat. Whereas Venter produces partial sequences of genes of unknown function to produce structural maps of the human genome, Matsubara wants to make a 'body map' of human genes that will show where they are expressed in the body and their function.

Leaders of other major genome projects in Japan are taking a similar position to Matsubara. Yuzo Minobe, director of Japan's rice genome project at the National Institute of Agrobiological Resources in Tsukuba, says that all results from the multi-million dollar effort to sequence the rice genome will be made freely available and that no patent applications should be filed for rice DNA sequences.

Takashi Yura of Kyoto University, who leads a project to sequence the complete genome of the bacterium *Escherichia coli*, an organism widely used in the biotechnology industry, says he has "absolutely no intention" of filing patents. His group is preparing to publish a 110-kilobase sequence, and it is sequencing another 250 kilobases out of the total genome of 4,700 kilobases.

There have been rumours in the United States that Eichi Soeda of the Institute of Physical and Chemical Research (RIKEN), whose team is sequencing chromosome 6 of yeast, is planning to patent yeast DNA sequences. But Soeda says he has no intention of doing so with the industrially important organism. **David Swinbanks**

Betting on rice

Tokyo

JAPAN'S effort to map and sequence the rice genome is moving forward rapidly, thanks largely to a craze among Japanese 'office ladies' for betting on horses. The huge boost in funds will raise its budget to a level comparable in scale to Japan's human genome project.

Last year, the Japanese Racing Association (JRA) — an organization funded by racehorse gambling — decided to put money into Japan's rice genome project (see *Nature* 353, 99; 1991). The association predicted it would contribute \$6 million a year on top of the \$2.7 million already being spent by the Ministry of Agriculture, Forestry and Fisheries. But in December, JRA boosted its contribution to ¥2,000 million (\$15 million), and the organizers of the project now expect the total annual budget, including government funds, to be at least ¥2,400 million yen (\$18.5 million).

The association donates about 25 per cent of its earnings to government projects, including research. It is flush at the moment because racehorse betting is particularly popular among young office workers. Similar associations for motor boat and bicycle racing also give money to Japanese science, but these sports are not as popular among young Japanese women.

Project organizers expect the current 40-member team from industry and the National Institute of Agrobiological Resources to more than double in size over the next several years. The seven-year project is backed by a newly formed association of 160 companies called the Society for Techno-innovation of Agriculture, Forestry and Fisheries (STAFF). Six companies — Mitsui Toatsu Chemicals, Mitsubishi, Kirin, Japan Tobacco, Sanyo Electric, and Hitachi Chemical — are providing researchers for genome analysis and technology development.

The first few years of the project will develop physical and linkage maps of the genome as well as a cDNA catalogue. Researchers will move onto mapping chromosomes and isolating useful genes. **D.S.**

MITI makes its move, cautiously

JAPAN'S powerful Ministry of International Trade and Industry (MITI) is at last taking its first step into genome research.

In a few weeks, the biochemical industry division within MITI will form a committee to coordinate various small projects that involve DNA analysis and to consider a project to sequence the genomes of industrially useful microorganisms. The eventual aim is a permanent government centre for DNA analysis.

MITI has hesitated to join genome research partly because the topic lies in the domain of other ministries but also because industry has shown little interest. And Masahiro Hashimoto, deputy director of the biochemical division, emphasizes that the present initiative is meant to strengthen DNA analysis and is not a genome project. Hashimoto also admits

that it may take a "long time" to persuade the Ministry of Finance that such a centre is needed.

The committee will try to coordinate three small MITI projects under the jurisdiction of the biochemical industry division: a project within Japan's huge fifth-generation computer project that is developing computer systems to handle the vast amounts of data arising from genome research (see *Nature* 345, 467; 1990); the application of nanotechnology, such as the scanning tunnelling microscope, to an analysis of DNA at a new interdisciplinary research centre in Tsukuba (see *Nature* 351, 90; 1991); and a small project within MITI's huge global environment research programme to isolate and develop photosynthetic microorganisms to absorb carbon dioxide. **D.S.**

Small, quick grants proposed as lifeline

Washington

US SCIENTISTS, believing that there is no time to waste, last week urged their government to fund hundreds of small collaborations with their Russian counterparts with part of \$400 million set aside for the destruction of nuclear weapons in the former Soviet Union. Under the plan, conceived during a two-day workshop held earlier this month at the National Academy of Sciences (NAS), those scientists who already hold federal research grants would identify partners in the former Soviet Union and provide them with money for vitally needed materials and equipment, visits to the United States, and collaborative research.

"If we wait too long, they'll be dead", says one participant, F. Albert Cotton, a professor of chemistry at Texas A&M University. "I'd hope that we could do something by May or June, certainly no later than six months from now."

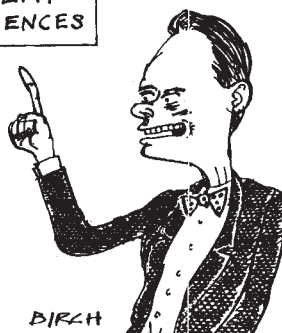
The proposal would keep active in the former Soviet Union those civilian scientists whose ability to carry out research has been severely curtailed by the political and economic chaos in the various republics. The goal is both to keep Russian scientists productive and to strengthen US science by tapping into a rich new source of available talent. The programme would draw as much as \$10 million from a fund created last fall by the US Congress to help the former Soviet Union to dismantle its nuclear arsenal and to employ those who built those weapons. The Bush Administration has already pledged to spend \$25

million from that fund on an international science centre (see *Nature* 355, 756; 1992).

The latest proposal is contained in a report issued last week by a group of 120 scientists and science administrators from government, business and the academic community who met on 2-3 March under the auspices of the NAS. The 23-page report, sent to D. Allan Bromley, the presi-

I propose an automatic grant to any ex-Soviet scientist who personally demolishes a nuclear weapon.

NATIONAL ACADEMY OF SCIENCES



dent's science adviser, also asks the US government to spend \$5-10 million on large-scale bilateral research projects and \$25 million to foster technology with commercial potential, as well as helping to create a fund of \$50-\$100 million to preserve the most important research labora-

tories in the former Soviet Union. It also urges the administration to adopt a proposal by US Representative George Brown (Democrat, California) for a binational foundation, endowed with \$200 million from public and private sources, to support joint research projects.

Bromley invited the academy to tackle the scientific crisis in the former Soviet Union after Boris Yeltsin, president of the Russian Republic, asked President George Bush for technological assistance during a meeting between the two leaders in late January. Bromley and Frank Press, president of the NAS, discussed the report during a hearing earlier this week before the Foreign Relations Committee of the US Senate.

There was no consensus at the academy meeting on how quickly such collaborations should be set up, nor on the best way to arrange them. Several participants favoured the approach used by a programme already under way at the US National Institutes of Health (see story below), which follows normal NIH procedures but allows for a slightly abridged proposal and a somewhat more rapid review. Other scientists preferred a radically shortened procedure that relies on the judgement of the grant holder, requires no more than a one-page description of the collaborator and the collaboration, and takes only a few weeks. The latter is what David Pines, a physicist at the University of Illinois and a participant at the workshop, suggested in a letter sent in December to Bromley and Walter Massey, direc-

"Our motivation is to do good science"

Seven US scientists received grants this month from the US National Institutes of Health (NIH) to work with their colleagues in the former Soviet Union. But NIH officials believe strongly that not even a scientific crisis is reason enough to abandon their system of picking the most worthy proposals.

While many of their colleagues are urging the US government to act quickly (see main story), these scientists say that the old ways are also the best. Paul Neiman, who heads the basic science division at the Fred Hutchinson Cancer Center in Seattle, Washington, has just added Victor Lobanekov and two associates at the All Union Cancer Research Center in Moscow to Neiman's existing NIH grant to study oncogenes in chickens. Neiman, who met Lobanekov in 1989, says that he would not collaborate with a colleague whom he did not already know well, and does not mind waiting four to six months to learn whether his research passes muster with NIH.

"It's great that we could help out Victor," Neiman says about his three-year, \$60,000 grant from NIH's Fogarty International Center. "But our motivation is to do good science, not to provide humanitarian aid."

Douglas Berg of Washington University in St Louis, Missouri, who will be working on the metabolic regulation of bacteria with Alexander Mironov and the eight members of his laboratory at the Institute of Genetics and the Selection of

Industrial Microorganisms in Moscow, agrees that haste might make waste. "Experiments have to be planned", he says, "and I think what we'll be doing is better because of the review."

Last year, the Fogarty Center began a programme for existing NIH grantholders who wanted to collaborate with colleagues in Latin America and the formerly communist countries of Eastern Europe. Last fall, the programme added countries from the former Soviet Union itself, and on 6 March it announced 12 such awards.

Those awards were part of a pool of 55 requests received last fall—a success rate that barely exceeds 20 per cent. The fairly detailed proposals were processed in the same way as any NIH grant application—a thorough review by a panel of their peers, and a second review by an advisory body to the center. Last month the center began to process a second round of 80 applications, including 26 that involve scientists from the new republics. Officials expect that a similar percentage will be successful after a final review is completed in May.

Philip Schambra, director of the Fogarty Center, says that he will resist any pressure either to speed up the process or to conduct a less thorough review. "We don't want to be open to criticism that we threw money away", says an assistant, Gray Handley. "You need more than one page to conduct a proper review, and to be able to justify your decision." J.D.M.

tor of the National Science Foundation (NSF).

"There are literally hundreds of people out there who, with a modest add-on to their existing NSF grant, could adopt research groups in the former Soviet Union", Pines says. "The money could be used to fly them here, bring them up to date, and arrange a mutually beneficial collaboration on a significant problem once they return to Russia. The point is to get a great number of people over here, as quickly as possible."

As with all recommendations that ask for more, the chief question is how to pay for it. Using money from the special fund set up last fall is attractive for two reasons: it protects the budgets of domestic research agencies, and it relies on money already appropriated.

But at a time when millions of US workers are unemployed and the federal deficit is approaching \$400,000 million, the notion of foreign aid is not a popular one. Recognizing that fact, those who attended the workshop hasten to explain that such a programme would benefit the United States as much as it helps the former Soviet Union.

"In order to have a free market in the former Soviet Union, which we want to foster, you need a strong science and technology base", explains Glenn Schweitzer, an expert on Central and Eastern Europe at the academy. "We also want to keep weapons scientists in place, working on civilian projects, instead of letting them go off and work for somebody that could threaten us. In addition, if we don't do it, the Japan and the Europeans will." **Jeffrey Mervis**

BIOTECHNOLOGY

'Silber's folly' goes public

BOSTON University (BU) will learn later this month what investors think of its controversial decision to invest nearly \$100 million over the past five years in a small biotechnology company.

The Hopkinton, Massachusetts, company, Seragen Inc., is preparing its first public offer of three million shares — 25 per cent of the company. If successful, the offer will raise enough money to support the company for the next 30 months and derive a product from its technology, which uses genetically engineered fusion toxins to kill disease-carrying cells. It will also enable the university to end its monthly infusion of \$1.2 million into the company, drawn from the school's endowment.

If the stock market responds as well as BU is predicting, university president John Silber will have proved wrong the critics of his 1987 decision to invest the equivalent of one-fifth of the university's endowment in the company. With shares priced at between \$17 and \$20 each, Seragen would be worth \$166 million on

Science cements Chinese ties

Jerusalem

ISRAEL has taken advantage of the high standing of its scientists in China to bring about full diplomatic relations between the two countries.

THE opening of a fully staffed embassy in Beijing in January ends a 40-year diplomatic silence between the two countries. Although in the mid-1980s China began to accept tourists visiting on group visas, scientists were the only Israelis permitted to enter the country on individual visas. Despite the official cold shoulder shown to most of their countrymen, these scientists were greeted warmly by their Chinese colleagues.

Reuven Merhav, an Israeli diplomat who had worked in Hong Kong, decided to put that affection to the test shortly after being appointed director-general of the Israeli Foreign Ministry in November 1988. He asked Joshua Jortner, president of the Israel Academy of Sciences, and its director, Meir Zadok, if the academy would like to establish an Israeli scientific centre in China. The academy officials quickly agreed, and in May 1989 they dispatched a three-person team to Beijing to sound out leading members of the Chinese scientific community.

The trip was a success. "They thought we were a scientific superpower," Zadok says with a laugh. In fact, Zeev Tadmor, a chemist who is now head of the Technion in Haifa, discovered that an undergraduate textbook he wrote with C.G. Gogos, *Principles of Polymer Processing*, had

been translated into Chinese and was being widely used.

A few days after the Israelis returned home, Chinese tanks rolled into Tiananmen Square to suppress the student uprising. Although China's use of military force against its own citizens invoked a strong reaction from the international scientific community, it proved to be a boon to the deepening Israeli-Chinese relationship. In fact, Israeli officials believe that China reacted with uncharacteristic speed to their political overtures in part to counteract their feelings of isolation brought on by sanctions imposed by much of the rest of the world.

As a result, what was expected to take years bore fruit in a matter of months. In November, Merhav and Zadok flew to New York to meet Li Lu Ye, the Chinese ambassador to the United Nations. They agreed that the Israeli academy would set up a liaison office in Beijing and that China would open a tourism office in Tel Aviv. The Israeli office — the first official Israeli presence in China — opened in April 1990. It was headed by Yossi Shalhevet, a 60-year-old agronomist and former director of the Volcani Research Institute.

"He's straightforward and he's warm," says Zadok about Shalhevet. "He's probably the best diplomat that Israel has."

Shalhevet says that the Chinese especially admire Israel's military prowess and its success in making the desert bloom. "They attribute our successes to the use of science and technology, which today is the highest priority in China," he says.

Once he arrived in Beijing, Shalhevet began to tour universities and scientific institutions throughout China. He paved the way for 100 Israeli scientists to visit China, and likewise for 50 Chinese scientists to tour Israel. He began a student exchange program and initiated research symposia.

Building upon his success, Merhav again met the Chinese ambassador to the UN in October 1990. He argued that the forthcoming peace talks on the Middle East required the posting of an experienced diplomat in its Chinese liaison office. The Chinese agreed, and last year an Israeli diplomat joined Shalhevet.

The multinational peace conference in Moscow last December further accelerated the process. Israel agreed to China's wish to participate in the talks only after China acceded to its request for full diplomatic relations.

Shalhevet remains in the Beijing office as scientific liaison. He says he is now busier than ever, bringing together business officials as well as arranging visits to Israel by Chinese scientists and students.

Abraham Rabinovich

Christopher Anderson

Physics losing the corporate struggle

SQUEEZED by economics and a corporate preoccupation with the bottom line, three of the largest US companies doing basic physics research are cutting back operations and reducing their research staff. AT&T's Bell Laboratories, Bellcore and IBM are all shifting their research programmes to applied research, a move that has already closed laboratories and left dozens of researchers looking for jobs. In time the move is expected to eliminate more than 300 full-time basic research positions.

At Bellcore, the research consortium begun nine years ago after a federal judge split the Bell Telephone monopoly into seven regional 'Baby Bells', basic research is near extinction. Of the 100 full-time PhD research positions now devoted to basic research at Bellcore, only about 50 will be left when the consortium completes its restructuring later this year. "We don't intend to continue any basic research as an isolated entity," says Murrae Bowden, assistant vice president for network technology research.

Superconductivity and nanofabrication research are being phased out, he says, part of a broader shift away from condensed-matter physics. In its place will be more software technology. "This industry is driven largely by software," says Bowden. "You have to ask yourself, how far down the chain of research do the owners want us to go?"

Bell Labs, traditionally the top of the heap in industrial basic physics research, is also moving away from its roots. "We are making a shift from the material and physical sciences to software research," says Kumar Patel, director of the division of materials science and engineering. After decades of steady profits from its monopoly ownership of the telecommunications market, AT&T is now fighting for its corporate life against stiff competition. And it is finding long-term research hard to justify.

So far the cuts have been modest, although corporate officials do not rule out the possibility that Bell Labs will one day bow out of basic research. Some of the eliminated programmes — such as the one in astrophysics — go back for more than 50 years. "But if we're going to pay Peter, we've got to rob Paul," explains Patel. "This kind of redirection is never easy, but I think that people understand the needs of the corporation."

Patel says that the balance between basic and applied research — 60:40 five years ago and 55:45 now — in five years will favour applied research, a transition that will eliminate about 100 basic research positions. Attrition, rather than layoffs, will cover the reductions. "Under normal circumstances," Patel says, "we'd

hire maybe 20 new people each year. Now maybe we'll hire 10."

Along with astrophysics, Bell Labs is phasing out most of its basic superconductivity research. Despite the Nobel Prize won by two Bell researchers — Robert Wilson and Arno Penzias, the current vice president for research — for discovering in 1965 the background radiation that offered evidence of the Big Bang, astrophysics "doesn't really couple with the needs of the company anymore," says Patel. And in superconductivity, Bell is dropping most of its research on the basic properties of the materials in favour of applied research on superconducting wires and the like. The one exception is superconductivity in C_{60} , the new-found form of carbon.

IBM, another traditional behemoth in basic physics research, is making a more gradual shift. Despite experiencing the first money-losing quarter in the company's history last year, IBM is not abandoning basic research, says James McGroddy, vice president in charge of research. Its corporate evolution has been in the making for the past ten years, he says, and is a transition from research at the bottom of the "food chain" to more work at the applied end. Nevertheless, although IBM's overall research budget has remained essentially flat over the past several years, the shift to applied research within that total is expected to mean the loss of about 100 basic research positions in the next few years.

Behind this widespread corporate retreat from basic research is the changing nature of the US electronics and telecommunications industry. In 1980, when AT&T and IBM were virtually unchallenged in their respective fields, such big companies maintained heavily endowed basic research laboratories almost as a matter of pride. AT&T, for example, was legally required to spend a percentage of its income on basic research in exchange for its monopoly position before the breakup. In a span of 40 years, such an emphasis produced the transistor and half a dozen Nobel prizewinners.

But AT&T must now compete for its market share, just like any other big company. For both the new Bell Labs and Bellcore, this has come as something of a shock. "The old Bell Labs had a key advantage — it couldn't fail," says Bellcore's Bowden. "We can."

After the breakup of the Bell system, Bellcore "came out of it with a mindset that our model of research was the Bell Labs model," Bowden says. That is, the best basic research and plenty of it. "We were like a kid in a candy store," he says. "We put too great an emphasis on materials research. It took some time to get it straight."

By the time the dust had settled, Bellcore's new owners were asking why they should be supporting a basic research programme when they are forbidden by law to manufacture products. The answer from scientists — that research would make the telephone companies a more "educated consumer" of high technology — was not persuasive.

Facing competition on all fronts, AT&T was also asking what its legendary research division had done lately. In 1989, the word came down that the old model of freewheeling and unencumbered basic science at Bell Labs would have to go. Now, says Patel, "we have a narrow view of what's important to us in the long run. What we call basic research is what fits the general needs of the company."

Researchers say these three companies may be only the most visible examples of the decline of basic industrial research. What is happening at Bell Labs, IBM and Bellcore "is not very different from what's going on in research labs throughout industry," says John Rowell, former director of Bellcore's research programme and now vice president for research at Conductus Inc., a superconductivity startup (see story on facing page). As the person who built up Bellcore's research programmes in the first place, Rowell is unhappy with the fact that companies with a combined annual income of \$90,000 million employ fewer than 100 basic researchers.

"One person in basic hardware research for every \$1 billion in revenue is a pathetic amount," says Rowell. "As far as the Bells go, anything that can not be immediately applied to telephone research isn't going to happen. And immediately means a year or less. To my mind that's not research."

Christopher Anderson

US SPACE PROGRAMME

Industrialist to NASA

DANIEL Goldin, an executive at TRW and a virtual unknown in the space community, was nominated last week to be the new head of the National Aeronautics and Space Administration (NASA). An engineer who now manages his company's work on military satellites and the Strategic Defence Initiative's 'Brilliant Pebbles' project, Goldin is seen as someone who would strongly support the Administration's space policy, including the planned Mars mission and the space station and a phase-out of the shuttle fleet. Critics saw the nomination as further evidence of NASA's domination by the National Space Council and its chair, vice-president Dan Quayle, and vow to oppose the nomination.

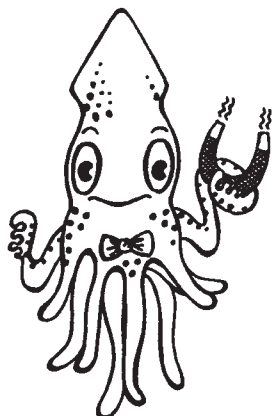
Christopher Anderson

SUPERCONDUCTIVITY

Growing up in public

WHILE the big corporate world seems to be losing interest in superconductivity (see adjacent story), Wall Street is just discovering it. Investors are jumping at their first opportunity to buy a share of some of the small companies that were created in the wake of the discovery of high-temperature superconductors in 1987.

Last December the first of those startup



Mr. SQUID

companies to go public — the Boston-based American Superconductor Corp. — saw its stock price double virtually overnight, from an initial offering price of \$11-\$13 to well over \$20, where it has held since. The sale brought in \$24 million, something that did not go unnoticed in the nascent industry. Later this month Conductus Inc., a five-year-old company based in Sunnyvale, California, will also go public, with an offering of 1.8 million shares at about \$12 each. Later this year, Superconductor Technologies Inc., based in Santa Barbara, California, also intends to start selling stock in itself.

Credit the mini-boom to pent-up investor interest, says Kevin Ott, executive director of the Washington-based Council on Superconductivity for American Competitiveness. "For five years you had investors hearing about superconductivity with nothing to buy," he says.

Even so, it may not be a stock for everyone: none of the young companies is close to going into the black. Most of the superconductor products the startup companies now sell are relatively modest, from sensors and wires to an educational package called "Mr SQUID" on which Conductus is pinning its hopes. "Investors are looking down the road here," believes Gregory Yurek, president of American Superconductor.

Nevertheless, Wall Street's response has been a pleasant surprise to the industry. "Everyone owes American Superconductor a debt of gratitude," says John LaVille, the chief financial officer for Superconductor Technologies.

Christopher Anderson

INDIA/RUSSIA COOPERATION

Joint projects lack money

New Delhi

A 1987 AGREEMENT for wide-ranging cooperation between India and the Soviet Union, predicted to last for 20 years, is on the brink of collapse as a result of the economic and political changes that have dismembered the former Soviet empire.

Next month, a delegation of Indian scientists will travel to Moscow to learn whether their Russian counterparts can afford to maintain a quarter-century of cooperation between the two countries. India has long depended on Soviet scientists for progress in many fields of science and engineering.

In 1987 the then leaders of the two countries, Mikhail Gorbachev and Rajiv Gandhi, agreed to some 85 joint projects in 14 areas, some new and others continuations. Today, not only have both men been removed from office — one by an assassin, the other by a revolution — but their successors have repudiated many of their policies.

The Russian Republic, which has taken on responsibility for roughly three-quarters of the projects covered by the agreement, is unlikely to support more than a small number of them, according to a spokesman for the republic. Anything not considered to be a "priority project", he says, could well be terminated after next month's meeting. "Other republics are free to participate if they also share the financial burden", he adds.

Indian officials say the future of the agreement "is difficult to predict". One-quarter of the institutions involved are located in the other independent republics, with financial difficulties equal to those facing the Russian Republic, and Indian scientists doubt that those governments will uphold the commitments made by their former rulers.

Their conclusions are based on the impact of political changes in the former

Soviet Union over the past few years. A powder metallurgy centre in Hyderabad has been delayed indefinitely as workers wait for the shipment of critical equipment. The Indian government is seeking a new source of the raw material for a facility in Uttar Pradesh that will produce a polio vaccine, having all but given up on its Russian suppliers. The Department of Atomic Energy has decided to complete a synchrotron radiation facility on its own after Soviet scientists failed to provide designs and some components that had been promised.

Other projects have been affected by an end to the movement of scientists between the two countries. What was once an annual flow of 1,000 scientists from Moscow has all but dried up. One culprit is the air fare — a one-way ticket from Moscow to Delhi costs 80,000 rubles and is unaffordable for the average Russian scientist. Especially affected are projects involving special lasers, high-temperature superconductors and engineering materials.

The 1987 agreement is not the only casualty of the current shift in the political landscape. Atomic energy officials have been waiting for months to learn whether the new rulers in Moscow will uphold a promise in 1989 to build two 1,000 MW nuclear reactors. The same is also true for five thermal and hydroelectric power projects. Gorbachev put his personal stamp on a joint project to build a hypersonic airplane, but work on the project has slowed to a crawl.

Russian officials say that the purpose of the meeting next month is to decide which projects are the most important. Indian officials, who are keen to continue with as many projects as possible, hope that the breakdown in bilateral scientific cooperation is temporary.

K.S. Jayaraman

Russians look to India for jobs

At the same time the Russian government is pulling back from its scientific collaboration with India, individual scientists are eager to improve those ties.

Four scientists from the former Soviet Union have applied for jobs in Indian laboratories. Their applications were sent to the Council of Scientific and Industrial Research and forwarded to the Department of Science and Technology.

Rama Rao, secretary of the department, says that he welcomes the enquiries as an opportunity for his country to acquire knowledge that could readily translate into scientific progress. "I hope that they do not have to leave", says Rao, "because their country needs them. But if they want to come to India, we will surely welcome them."

The Indian government has not acted on the requests because it has no policy on hiring foreigners to work in Indian laboratories. But Rao says that India is a natural choice for such emigrants because of their long-standing ties to the Indian research community.

K.S.J.

UV researchers opt for a compromise

Washington

WHEN researchers reported earlier this winter that ozone loss had reached record levels at both poles, the US public wanted to know one thing: Were levels of the most harmful ultraviolet radiation (UV-B) going up as the ozone layer became thinner?

Good question. But despite a UV monitoring network that has been in place for more than 20 years, researchers still have no good answer. The problem is that the existing monitors (known as Robertson-Berger meters) were designed to give a broad 'sunburn equivalent' reading. They are not sensitive enough to differentiate between UV-A radiation, which is believed to stimulate cell repair, from UV-B, which damages living tissue. Worse, most of the monitors are in urban areas, where the ozone that is pumped into the lower atmosphere from man-made pollutants often cuts out much of the UV radiation that penetrates the stratospheric ozone.

Last week, a group of researchers met in Washington, DC, to design a better system. Researchers sponsored by the US Department of Agriculture (USDA) have

developed a more sensitive UV monitor, and the agency wanted advice on just how and where the new monitors should be deployed. For one thing, the new monitors will be expensive (more than \$100,000 each at the moment) and they may not be available in quantity for several years. The agency hopes eventually to have at least six monitors in the continental United States as part of its \$3 million programme.

Deciding where to place them will not be easy. The first one will almost certainly join other monitors at a Department of Energy experimental station on the Kansas-Oklahoma border, where the UV reading can be correlated with other radiation and atmospheric measurement. Others could be in almost any rural area, however, and it is not clear just what kind of geographic location would give the most useful measurements. For example, researchers do not know if two locations just a few hundred miles apart would produce significantly different readings.

To answer some of these questions, the scientists at last week's meeting proposed something of a stopgap solution: deploy

10-15 meters similar to the old Robertson-Berger monitors in the sort of locations in which USDA would like to put the more sophisticated monitors. They are expected to make such a recommendation in a report that will go to USDA in early summer.

But some researchers are concerned that this approach does not tackle the problem head-on. They point out that other countries, including Canada, New Zealand and Australia, are taking good measurements with high-quality scanning spectral radiometers that are already available.

James Gibson, a Colorado State University researcher who led last week's conference, has heard this complaint before. He was involved in the 10-year, congressionally mandated acid rain assessment project that spent so long perfecting its research that legislation was passed before its final report was released in 1990. He says that the Robertson-Berger solution appears to be the best compromise between waiting for the best possible measurements and leaping in with cheap monitors now.

Christopher Anderson

SCIENTIFIC MISCONDUCT

FBI investigating NIH leaks

Washington

THE US Federal Bureau of Investigation (FBI) is looking into allegations that current and former employees of the National Institutes of Health's (NIH) misconduct office have leaked documents relating to the agency's investigation of scientific misconduct. The FBI's actions, in turn, have triggered a new investigation of NIH by Representative John Dingell (Democrat, Michigan).

For the benefit of those who are still keeping score, that makes it Congress investigating NIH's investigation of a former employee's investigation of NIH's investigations of misconduct. Ironically, by the time all this is over, the misconduct office is expected to be history (see page 191).

Last week, an FBI agent interviewed Suzanne Hadley, the former deputy director of the NIH Office of Scientific Integrity (OSI), who has since turned whistleblower and is working part-time with a congressional investigations subcommittee that Dingell chairs. According to Hadley, the agent told her that he was looking into allegations that "documents were walking out of OSI" and that the FBI had interviewed several OSI employees who had admitted giving documents to Hadley. The agent did not identify a specific crime that the FBI was investigating.

Although Hadley declined to answer questions without a lawyer present, she says that the FBI probe "is all innuendo. It

has the feel of a fishing expedition." She denies taking documents, although she acknowledges that she has been given documents that have been leaked. The same could be said, of course, of reporters, the Dingell committee and many scientists.

Last week NIH replaced the locks on Hadley's office door "so that nothing is changed", says spokeswoman Joanna Schneider. Hadley will be given another office at NIH, she says.

NIH also changed the locks at OSI last week — the third time in the past two months. Officials first changed OSI's locks after the *Chicago Tribune* reported on a draft of a final report in the case of NIH AIDS researcher Robert Gallo, the fourth such leak in two years. Only after the article appeared did it occur to OSI officials that they had not changed the locks last July, when Hadley was reassigned to another part of NIH. Hadley says that, since then, she has entered OSI only once (and with an official escort). OSI changed its locks a second time after receiving a letter from Dingell complaining that the office had shredded sensitive documents.

NIH officials say the FBI investigation began when the OSI received a call from a former employee alleging that current employees were providing documents to Hadley. NIH took the allegations first to the Inspector General of its parent agency, the Department of Health and Human Serv-

ices. But the Inspector General declined to investigate because of a possible conflict of interest with an ongoing investigation of NIH. That inquiry is looking into allegations that NIH officials made false statements in connection with the 1984 US patent application for a blood test for the human immunodeficiency virus, which is based on the work of Gallo. The Inspector General then referred the case to the FBI. According to Hadley, the FBI agent said that NIH has determined that the leaks and their political fallout have paralysed OSI, rendering it "dysfunctional" and costing the agency \$9,000 per day.

Hadley alleges that the FBI investigation is just the latest example of a "harassment campaign" that NIH launched after Bernadine Healy, the NIH director, criticized her "lack of objectivity" in handling the misconduct case of Tufts University researcher Thereza Imanishi-Kari. This week, Hadley's attorney, Wayne Schrader of Gibson, Dunn and Crutcher, expects to file a complaint with the Office of Special Council (an independent US agency) alleging that NIH have violated the Whistleblower Protection Act in essentially demoting Hadley and making false statements about her. In a letter dated 12 March to Healy, Dingell called NIH's action an "apparent act of harassment and intimidation aimed at [a] courageous, public-spirited whistleblower[er]."

Christopher Anderson

Reforms win praise, but not patrons

London & Washington

The United Nations Educational Scientific and Cultural Organization (UNESCO) has begun to restore its tarnished reputation. But it appears unlikely that even a spate of reports attesting to its improved efficiency and professionalism can regain its lost patrons, in particular the United States and Britain, and accomplish the goals set by its director-general, Spanish biochemist Federico Mayor.

Two weeks ago, a report compiled by Knut Hammarskjöld, formerly head of the International Air Transport Association, and Peter Wilenski, Australia's permanent representative at the United Nations, praised Mayor for removing much of the waste and cronyism that characterized UNESCO under the 13-year reign of his Senegalese predecessor, Amadou Mahtar M'Bow, that ended in 1987. Next month a similarly favourable assessment is expected both in an internal report by the US Department of State and a review by the US congressional General Accounting Office, which sent a nine-person team to comb through the Paris-based agency's finances.

UNESCO officials hope that these reports will help to convince the United States and Britain, which walked out of UNESCO in 1984 and 1985, that the time is now ripe to rejoin the organization. With the former Soviet republics so far unable to take over the Soviet Union's \$30 million-a-year UNESCO subscription, the infusion of cash that the two countries would bring would be an important boost for the agency. But, for the time being at least, it seems that UNESCO will have to live with its cash-flow problem.

British officials agree that UNESCO seems to have put its worst excesses behind it, but they want to see Mayor's reforms take deeper root before agreeing to rejoin. And in the United States, a Democratic party victory in the forthcoming presidential election seems the only hope for UNESCO's proponents. That is because President George Bush remains vehemently opposed to rejoin-

ing an agency that costs the United States more than \$50 million a year.

In the 1980s, UNESCO was widely viewed as overstaffed, dependent on the services of expensive, outside consultants and lacking in any mechanisms to appoint and promote staff according to their merit. For Britain and the United States, the final straw was M'Bow's support for a programme called the New World Information and Communication Order. The new world order was intended to help developing countries to manage the flow of information to their people, but many people saw it as condoning government censorship of the media.

Hammarskjöld and Wilenski now say that Mayor has successfully implemented a new merit system for recruiting and appraising staff. In a move that is almost unheard of within the United Nations system, according to UNESCO's deputy director-general, Chatan Lal Sharma, some of the worst "non-performers" have been dismissed. Although only four staff have actually been fired, Sharma believes that other offenders within UNESCO will now quickly mend their ways. UNESCO's controversial information and communications policies are also now believed to be substantially reformed.

Most of the reforms has been carried out since 1990, when a US State Department report charged that Mayor had made little headway in imposing changes on a staff still dominated by people appointed by M'Bow (see *Nature* 344, 801; 1990). And the Hammarskjöld and Wilenski report warns that the progress could "easily regress" without strong backing from UNESCO's member states.

"It's like losing weight," says Sharma — early progress can be easily wiped out by a lapse in self-control. He says that it may take another two or three years for Mayor's reforms to become "institutionalized", with the first task being to introduce a promised system to award able staff with rapid promotion.

UNESCO staff accept that the recent reforms will not win over the current US Administration, despite support for the

agency from public figures such as Frank Press, president of the National Academy of Sciences. But there is a mood of optimism within UNESCO about the prospects for a British change of heart.

British officials say that they are also pleased with recent progress. In particular, they are encouraged by a plan to give a subpanel of UNESCO's policymaking executive board a greater role in overseeing the agency's finances and administration.

Starting next year, the full executive board will be made up of representatives accountable directly to their individual governments. At present, says one British official, UNESCO's policy is decided by a collection of "throbbing great brains". Although they are appointed because of their intellectual credentials, they face little pressure to steer the agency in the direction desired by its member states.

Nevertheless, Britain intends to take a 'wait-and-see' attitude before making a decision on UNESCO (although, as in the United States, a new government could bring a change of policy). British officials want more evidence that the member states will spend UNESCO's tight budget on a smaller number of priority areas, such as basic education in developing countries and scientific programmes related to global change. Although some progress has been made, there is general agreement that UNESCO's resources are still spread too thinly. Achieving the correct balance will require many of UNESCO's member states to accept that their own pet projects do not necessarily deserve money from UNESCO's budget.

The agency's two-year budget for 1992 and 1993 is less than that of many US universities, totalling only \$443 million. More than half comes from external sources such as the UN Development Programme, rather than subscriptions from member states. With Russia, Ukraine and Belarus having taken over the former Soviet Union's UNESCO membership, but not so far its subscription, even the \$15 million a year that Britain would be charged for membership would provide an important fillip for the agency.

Sharma plays down UNESCO's cash-flow problems, saying that many UN agencies face the same difficulty. In any case, he says, UNESCO is able to borrow more than \$60 million if it needs the money. But given the uncertainty surrounding the economic future of the former Soviet republics, the wisdom of taking such a course must be questioned.

**Peter Aldhous &
Christopher Anderson**

OCEANOGRAPHY

Explosion kills two

A DEPTH-CHARGE used in a secret US Navy oceanographic experiment exploded last week aboard a research ship off the coast of Washington State, killing two crew members. Although the Navy would not disclose the nature of the classified research, the operations were being conducted by a team of researchers from Johns Hopkins University's Applied Research Laboratory (APL). The laboratory con-

ducts seismic and ocean-floor mapping experiments for the Navy using the *Amy Chouest*, a chartered 265-foot research vessel.

None of the APL researchers was injured in the explosion, which occurred when the ship was about 13 miles offshore. The cause of the explosion has not yet been identified.

Christopher Anderson

Italian promotions defended . . .

SIR — The letter by G.F. Gaetani and A. M. Ferraris (*Nature* 353, 10; 1991) engendered a violent press campaign by some Italian newspapers, following which the Minister of University and Scientific Research and Technology ordered an inquiry into the proceedings of the committee and its decisions regarding the professorship in question.

As members of that committee, we feel it is now time to provide further information, as the inquiry has now been concluded and has given full approval to both our proceedings and our decisions.

We shall not comment upon the methods used by Gaetani — one of the “losers”, as he says — to influence his own academic promotion, neither shall we comment upon the evaluation of the different applicants. But we wish to make it clear that a competition for appointment as a university professor is quite different from academic competitions for scientific awards.

Research work and results are undoubtedly very important in evaluating candidates for professorial positions in a school of medicine. But as regards clinical teaching in haematology, the activity of an applicant needs to be evaluated also in terms of clinical and professional ability, as the basic duty and responsibility of a professor is to teach students how to diagnose and treat a disease and how to care for patients. Such fundamental aspects of the problem should not be ignored.

MICHELE BACCARANI* (Professor of Haematology Udine University), GIANLUIGI CASTOLDI (Professor of Haematology, Ferrara University), GIUSEPPE PAPA (Professor of Haematology, Rome University (Tor Vergata)), GIUSEPPE PERONA (Professor of Haematology, Verona University) SANTE TURA (Professor of Haematology, Bologna University)

. . . and attacked

SIR — Promotion to a full professorship in paediatrics in Italian medical schools seems to suffer from more serious problems than does a similar appointment in haematology¹. A committee of five full professors of paediatrics, after three years of discussion and the replacement of two of its members, has just appointed 25 new professors from among 74 candidates².

The explanation put forward by Gaetani and Ferraris¹ for haematology candidates seems to apply in an even more extreme sense here. Of the paediatrics candidates, five of the winners averaged 2.2 winners in the *Science Citation Index* (two having published no papers in international journals) whereas five of the

losers averaged 103, with a remarkable peak of 302.

The Minister for University Affairs decided not to validate these results and the same committee was asked to reconsider the same list of candidates. After two months, the results were confirmed. After a further three months, the National University Council (CUN) is to confirm the appointment of 24 new professors, casting doubts on only one of the five candidates with few citations. This means that the minister must now either accept the decision on the 24 winners or appoint a completely new committee to reconsider all 74 candidates.

It is not our intention to criticize the Italian system of academic promotion in general; it provides good results in many fields of pure science, including physics, where Italian work is well established internationally. But although scientific output should not, of course, be the only criterion for academic promotion, there should be a threshold below which no candidate is acceptable.

The case of the paediatrics professorships is not an isolated one, and may indicate systematic nepotism in the management of appointments to clinical professorships in Italian medical schools: candidates for paediatrics professorships, ten are *allievi* (pupils is not an adequate translation) of the five members of the committee, and all of them were appointed.

The main problem is that decision-making committees and university officials are not held responsible for their mistakes. Substantial portions of research grants and all new academic posts are distributed among universities, faculties and departments without consideration for the quality of their science or their teaching. This is not the situation elsewhere in Europe or in the United States. As a consequence, Italian departments and institutes have no incentive for high standards of science, and the appointment of the best scientists and scholars is not encouraged.

We should like in conclusion to suggest that the Italian system of academic promotion be changed to follow more closely the criteria used in other European universities. The departments concerned should be responsible for specifying the scientific and academic requirements for a vacant position and should be allowed to nominate at least three candidates. All the candidates, who need not be Italian, should submit their *curricula vitae* and reprints of their best publications to an international panel of referees. The committee should then say (1) whether at least three candidates meet the scientific and academic threshold qualification and

(2) who among the candidates best meets the department's requirements.

FERNANDO AIUTI* (professor of clinical immunology and allergy, University of Rome 'La Sapienza'), CARLO BARONI (professor of pathology, University of Rome 'La Sapienza'), ANTONIO CAO (professor of paediatrics, University of Cagliari), ANTONIO FANTONI (professor of molecular genetics, University of Rome 'La Sapienza')

1. Gaetani, G. F. & Ferraris, A. M. *Nature* 353, 10 (1991).

2. Aiuti, F., Baroni, C., Cao, A. & Fantoni, A. *Lancet* ii, 1387 (1991).

*Full address: Policlinico Umberto 1, 00161 Rome, Italy

SIR — Further to your continuing correspondence on Italian academic appointments, I wish to describe what seems to be an entirely novel way of constituting the commissions that evaluate applicants for tenured academic positions. My story also illustrates the remarkable flexibility now required of those eventually appointed.

Some time ago, the physics department of this small university in central Italy announced a vacancy for an associate professor in atmospheric physics; we have an active group in stratospheric modelling and measurement involved in several international research programmes. The several active workers in this field are concentrated at three or four universities, but in the whole of Italy there are only one full professor and one associate professor of atmospheric physics and one associate professor of meteorology.

On this occasion, an entirely novel way of choosing the evaluation commission was devised. For these purposes, all disciplines are grouped together on the basis of some kind of affinity; the subgroup including atmospheric physics (D043) is entitled Atmospheric Physics, Oceanography and Navigation, and belongs to the major grouping of Earth Sciences (not physics), including such subgroups as General and Spherical Astronomy, Terrestrial Physics and Limnology.

The evaluation commission was chosen in two steps; first, a pool of potential members was chosen by lot from among all professors and associate professors in the Earth sciences, after which the same people voted to elect a subset from the pool. It is useless to speculate about who invented this system, but the effect is evidently to favour the large schools, in this case professors of terrestrial physics, who happen to be concentrated in a few universities.

Even if the lottery were above board (which nobody believes), the result would have been that candidates would probably have been from the same large schools, whose votes then ensure the

election of the same people.

Proofs of this positive feedback are easily provided. Of 18 candidate full professors, seven were from the same southern university (Naples) and 13 were professors of terrestrial physics. It is relevant that terrestrial physics professors in Italy are mostly in solid-Earth geophysics, especially in seismology and volcanology (which explains why they are concentrated in southern Italy). Some years ago, a very good candidate for a full professorship in terrestrial physics was rejected on the grounds that his research articles (more than 40, in international refereed journals) in atmospheric physics were "not pertinent".

In the recent competition in the D043 subgroup, there were three vacancies to be filled — our own post in atmospheric physics, one in oceanography and limnology from another university and one at the Institute of Navigation of the University of Naples. The commission formed by lottery and election consisted of three full professors and one associate from the University of Naples and one associate from the other university with a vacancy to fill. None of them has ever published a paper on atmospheric physics, meteorology or the like. Two separate protests on this score by the president of the National Institute of Geophysics were, as usual, ignored by the Ministry of University Research and Technology.

What other countries would expect from such a commission I do not know, but in Italy we quickly saw an obvious flaw; this university, which had prompted the competition by announcing a vacancy, would be in trouble. Indeed, soon after the commission was formed, we knew the name of the seismologist who would win the position as associate professor of atmospheric physics.

In this case, the commission had to decide on the basis of publications, other evidence and a public lecture on a topic of the candidate's choice. Among the 12 applicants considered at the last stage were three in atmospheric science, one of whom had spent ten years at an international institution in Europe, and had acquired an international reputation for his work there. Six weeks after the last of the public lectures, the commission announced its decision. Of the three vacancies, the two at the other universities were filled by local candidates, but the vacancy in atmospheric physics at this university was filled by a seismologist (whose name we knew from the beginning) from the same southern university providing four out of the five commissioners.

Formally, the university has the right to decline to appoint the nominee, on the grounds that he is not what it was looking for, but the ministry has the right to send him anyway. There is also a slight chance

that the Commission on Universities might intervene, but it is empowered to do so only if there have been formal irregularities; provided that there have been no obvious irregularities, a university could have asked for a cardiologist and been sent a dentist instead.

Atmospheric physics is now growing in Italy in response to questions such as that of global warming, but atmospheric scientists are in danger of becoming an endangered species. Yet the ministry and the Commission on Universities have recently decided that atmospheric physics should be an essential part of the physics curriculum. It is doubtful that this goal can be achieved if tenured appointments are decided by power and not competence.

GUIDO VISCONTI

*Universita degli Studi L'Aquila,
Atmospherics Group,
67010 Coppito (L'Aquila),
Italy*

Drug approvals

SIR — The survey by the US Pharmaceutical Manufacturers Association (PMA) of new biotechnology products that have been approved or are in the pipeline (see *Nature* 353, 591; 1991) is interesting, but we at the Food and Drug Administration (FDA) were disturbed by its interpretation of the projected rate of marketing approval for new biotechnology products by the FDA's Center for Biologics Research and Review (CBER).

PMA has calculated FDA's rate of approval for new biotechnology products by dividing the total number of products approved by the total time elapsed since the first product was approved (in 1982), which yields a value of some 1.6 approvals a year. They then reason that because their survey reports 21 products now awaiting approval (although FDA's records indicate the number is closer to 30), it would take FDA's CBRR 13 years to approve these products.

The fallacy of this mathematical sleight-of-hand is that it ignores the fact that for much of the past decade there were very few products in the pipeline and that they were approved rapidly after the marketing applications were received by FDA. Moreover, at least 4 of the 21 are currently ineligible for approval because of marketing exclusivity granted to other similar products under the provisions of the US Orphan Drug Development Act.

Finally, PMA's own data reveal that "FDA required an average review time of 21.4 months" to approve biotechnology therapeutics compared to the mean approval time for other new molecular entities of 31.8 months cited in PMA's "New Drug Approval in 1990". Thus,

FDA has, in fact, approved new biotech products an average of ten months more quickly than for other non-biotech drugs, with all the advantages to manufacturers — and, more important, to patients — that implies.

HENRY I. MILLER
(Director)

*Office of Biotechnology,
Food and Drug Administration,
Rockville, Maryland 20837, USA*

Artists' offspring

SIR — The possibility that artists have more sons than daughters is suggested by a count of the following sources:

- (1) those for whom the relevant information is included in *Who's Who in Art*, 24th edition¹ (which covers artists, designers, craftsmen, critics, writers, teachers and curators);
- (2) those mentioned in the obituary section of the above publication for whom the relevant information can be found in *Who's Who in Art*, 19th edition²;
- (3) those artists and craftsmen in the 1985 to 1990 volume of *The Annual Obituary*³ for whom relevant details can be found therein. (Sons and daughters of artists in both (2) and (3) are counted once only.)

The total numbers of sons and daughters, including deceased children but excluding stepchildren, are 1,834 and 1,640 respectively, giving a sex ratio of 1.1183. A control experiment was carried out involving a count of the sons and daughters recorded in the general *Who's Who*⁴, the aim being to expose any male (or female) bias in reporting offspring: the sex ratio for the first 4,002 offspring was 2,046/1,956 = 1.0460. This is close to the 1.0534 ratio which applies to British births generally (based on statistics of 723,093 births published in the 1987 *Demographic Yearbook*¹) and suggests that there may be no significant male bias in reporting offspring in biographical reference books. Comparing then the sex ratio of the artists' children in (1) – (3) against the sex ratio for the British births generally, the χ^2 value is 3.0173, giving a probability $P=0.085$ approximately. Further work on correlations between parentage and the sex ratio of offspring appears desirable.

R. A. BECK

*49 Curzon Avenue,
Stanmore,
Middlesex HA7 2AL, UK*

1. *Who's Who in Art* (24th edn) (Art Trade Press, Havant, Hants, 1990).
2. *Who's Who in Art* (19th edn) (Art Trade Press, Havant, Hants, 1980).
3. *The Annual Obituary* Vols 1985–88 (ed. Burgess, P.); Vols 1989–90 (ed. Andrews D.) (St James Press, Chicago and London).
4. *Who's Who 1990* (Black, London).
5. 1987 *Demographic Yearbook*, 350 (United Nations, New York 1989).

Uncertainty of detecting sea change

Carlos M. Duarte, Just Cebrián and Núria Marbà

The alarming termination rate of long-term monitoring programmes in Europe is hindering the detection of ecosystem change in the ocean. Existing programmes must be linked and data shared.

GROWING concern about human influence on marine ecosystems conflicts with our inability to separate man-made from 'natural' change. This limitation results from the lack of adequate baselines and uncertainty as to whether observed changes are local or on a broad scale. Long-term monitoring programmes should be able to solve both these deficiencies, but failure to realize their potential has led to a shocking number of them being prematurely stopped, often permanently.

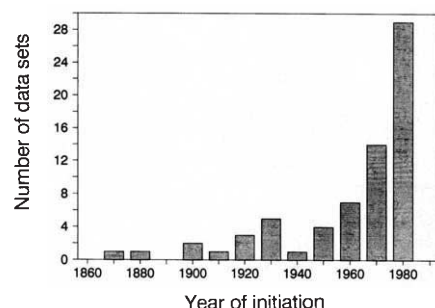
The initiation of new long-term programmes has been increasing exponentially since 1940 (see figure). Many (55%) of these programmes, however, have been stopped already, most often in the late 1980s when 40% were terminated. Consequently, few records of change in the sea extend beyond two decades, and efforts to monitor marine ecosystems in Europe are not increasing despite growing evidence that human influence is having a significant effect. Thus, long-term monitoring programmes are, paradoxically, among the shortest projects in marine sciences: many are initiated, but few survive a decade.

The growing number of new long-term monitoring programmes parallels a similar bloom in human damage to marine ecosystems. Funding of monitoring programmes triggered by these catastrophes are often withdrawn once 'recovery' has occurred, or once the issue is no longer attractive to communication media. Recent examples include winding down of operations to clean-up the oil spill during the Gulf War, the largest oil spill in history, because countries donating funds to the International Maritime Organization's Persian Gulf Oil Pollution Disaster Fund were no longer willing to contribute, even though clean-up and evaluation of ecological damage is far from complete¹.

The sustained support required by long-term monitoring programmes is jeopardized by unexpected drops in research funding caused by policy changes, such as that experienced recently in France². This appears to be the main cause of the alarming rate at which long-term marine monitoring programmes were stopped in Europe, many of these then cancelled during the 1986–87 funding crisis experienced in France and the United Kingdom^{2,3}. Long-term monitoring programmes are easy victims

of funding crises because, by their very nature, they are not competitive when evaluated from their short-term scientific yield. The frequent imbalance between the effort invested and the scientific yield, along with the routine nature of long-term monitoring programmes, does little to make them appealing to scientists looking for creative research to develop solid curricula. Consequently, continuation of long-term monitoring programmes is often heavily dependent on the personal effort and dedication of individual scientists.

The question of whether the marine environment should be monitored is no



Rate of initiation of long-term monitoring programmes in European marine stations.

longer at stake, and there is an explicit demand from an increasingly concerned society for answers as to how and how much the marine environment is changing. Scientists must, therefore, improve the mostly meagre scientific yield of their effort. The ability of long-term monitoring programmes to detect and characterize change in the marine environment will be best realized if data sets were examined in concert, as opposed to the isolation of present practices, thereby comparing observations from distant locations. A comparative approach is scientifically more sound because spatial and temporal scales of change in the sea are linked by mixing, such that processes responsible for decade-scale changes in the marine environment operate at spatial scales of thousands of kilometres⁴. Departure of ecosystem properties from long-term average values are therefore likely to be part of changes in water mass distribution and biogeographical patterns at oceanic scales. For instance, we observed anomalously high sea temperatures throughout Europe between 1988 and 1990, and large fluctuations in fish

catches can result from changes in the biogeographical distribution, rather than the size, of the populations⁵. Detection of these broad-scale changes, and distinguishing them from local changes, is possible only when data from distant locations are compared.

A comparative approach to long-term monitoring should yield valuable knowledge on the nature of changes in the sea and enhance the potential to forecast and detect them. It should encourage the solidarity and cohesion among individual programmes needed to reduce the presently alarming rate at which monitoring programmes are being stopped, because termination of individual programmes would impinge on the overall potential to establish broad-scale patterns of change.

The broad-scale, comparative approach to long-term monitoring of the marine environment proposed here must involve robust international management and funding systems, either already existing (such as EC, ICES or UNESCO) or new, in which individual efforts must be integrated to ensure continuity. Yet an effective approach requires that scientists involved in long-term monitoring programmes be prepared to share data in the relatively short term. Reluctance to share data is a serious threat to large-scale cooperative multinational efforts in other fields where there is a difficult balance between the rapid data dissemination required and the need for a period of confidentiality before the data are available to competing research groups. The balance may be particularly difficult to achieve in the case of marine monitoring programmes because the period of confidentiality of data sets is extremely long (generally more than 5 years). Yet we believe that the profits of a comparative, multinational approach should suffice to overcome any reluctance and improve the prospect for the detection of change in the sea. □

Carlos M. Duarte, Just Cebrián and Núria Marbà are at the Centre d'Estudis Avançats de Blanes, CSIC, Camí de Santa Barbara s/n, 17300 Blanes, Girona, Spain.

1. *Mar. Pollution Bull.* **22**, 480 (1991).

2. *Nature* **346**, 125 (1990).

3. *Nature* **326**, 237 (1987).

4. Harris, G. P. *Can. J. Fish. Aquat. Sci.* **37**, 877–900 (1980).

5. Hovgaard, H. & Burch, E. in *Large marine ecosystems. Patterns, Processes, and Yields* (eds Sherman, K. et al.) 37–43 (AAAS, Washington, DC, 1991).

Change in fraud review proposed

A strict definition of fraud and a plan to move the NIH Office of Scientific Integrity out of NIH are a step in the right direction of reforming a system that has not worked well.

FOR more than a decade, the biomedical research enterprise in the United States has struggled, without evident success, to develop fair and efficient procedures for dealing with allegations of scientific fraud. Universities have repeatedly claimed inexperience as an excuse for inadequate investigations, and the three-year-old Office of Scientific Integrity (OSI) at the National Institutes of Health has proved itself uniquely unable to cope. There may be light at the end of the proverbial tunnel.

A senior federal advisory committee has recommended that the government adopt a strict-constructionist definition of fraud and has supported a proposal that OSI be abandoned altogether—good advice on both counts.

One of the main roadblocks in fraud investigations has been the research community's rather astonishing inability to agree on a clear definition (*Nature* 352, 563; 1991). As a result those responsible for investigating allegations of wrong doing (including the fraud cops at OSI) have included in their purview not only outright data fabrication but also 'practices that seriously deviate from those commonly accepted within the scientific community for proposing, conducting or reporting research'. The trouble here is that there is no consensus about what constitutes commonly accepted practices, thus leaving it to the investigators to decide as they go along whether or not they approve of various laboratory practices.

The tangible consequence of this approach is evident in the case of the OSI v. Robert Gallo at NIH. Gallo who has been cleared of allegations that he 'misappropriated' an AIDS virus from French colleagues but he has nonetheless been charged by OSI with not running a tight ship.

The recommendations of the advisory committee, which reports to the Secretary of Health and Human Services (HHS), whose department includes NIH and other major biomedical research agencies, would go a long way to clarifying the issues. First, the committee would define fraud for what it is. Soundly rejecting the view often expressed at academic gatherings that fraud is difficult to define, the committee, chaired by Nicholas Steneck of the University of Michigan, has proposed defining fraud the way the dictionary does. If the Secretary takes the advice (as he should), research fraud would henceforth be defined as 'plagiarism, the fabrication or intentional falsification of data, research procedures or

data analysis, or other deliberate misrepresentations in proposing, conducting, reporting or reviewing research'.

It is significant that the committee has adopted this common sense definition that includes intent as an essential element in fraud. In so doing, Steneck and his colleagues have taken a position that runs counter to the current OSI position that intent is irrelevant in determining whether someone is guilty of committing fraud—a position with which NIH director Bernardine Healy also takes strong exception.

Another important recommendation for handling fraud—this one from NIH and HHS staff—while organizational in nature, would bring due process and a sense of fairness and impartiality to the investigatory enterprise. The OSI, now an integral part of the NIH with offices on the NIH campus, would be abolished, to be replaced by a new entity with the ungainly name Office of Research Integrity Assurance (ORIA) that would operate out of the office of the Secretary—along with the Office of the Inspector General, which enforces the department's laws. (It might even be wise to take OSI out of HHS altogether.) Under this scheme, the NIH would be relegated to a role in setting policy (the NIH director would head a new Research Integrity Policy Board).

The Steneck committee also tackled head on numerous complaints that current OSI procedures deny the accused a fair trial. Because OSI staff have taken the position that theirs is a search for scientific truth (rather than for evil-doers), they deny the accused access to the evidence and prohibit the cross-examination of witnesses. Under proposed reforms, the accused would always be entitled to appeal a finding of guilt to an adjudication panel whose activities would most likely be open to public view. They should be, just like the criminal trials they so closely resemble.

The principal argument against moving OSI out of NIH into the more scientifically impersonal reaches of the Secretary's office is that only scientists are qualified to investigate the work of other scientists. It is a valid thought, but the evidence does not bear out an inference that the present system works well. The Gallo case, and the case against Thereza Imanishi-Kari, who is accused of fabricating data in a paper she co-authored with David Baltimore, illustrate the point.

The OSI, understandably lacking the expertise in immunology requisite to an

analysis of the Imanishi-Kari paper, is being assisted by a panel of scientific experts, in theory much to the good. However, when two of the three dissented from some of the OSI staff's conclusions in the case (which were leaked to the press), OSI essentially ignored them.

In the Gallo case a panel of researchers from outside the NIH was named not only to provide scientific advice but also to serve as a watchdog, lest the OSI show favouritism toward Gallo as one of NIH's own (it has not). However, NIH to this day has not made plain just what it is the advisory panel, headed by Frederic Richards of Yale, is supposed to do, although it has been in operation for a couple of years. Furthermore, the committee (10 members strong) has itself been accused (whether justly or not) of bias against Gallo.

The OSI's efforts to get independent advice in these prominent cases have fallen short. Perhaps a revamped investigatory office could do better. In any case, a new fraud office, operating with a strict definition of fraud, could escape the illusion that resolving allegations requires establishing scientific truth. In the Imanishi-Kari case, Baltimore has argued that others in the field have produced data that tend to support hers. This is pertinent to a discussion about whether the hypothesis in the paper might be true; it is not relevant to the charge that her data were fabricated.

An additional matter of some importance in investigating fraud is confidentiality. So far, OSI has been unable (or unwilling) to keep its draft reports from finding their way to the press. NIH director Healy has become so distressed about leaks that she has taken the aggressive step of calling in the FBI to determine whether present or former OSI employees are guilty of criminal behaviour (see news story, page 186). The whole business of leaks is another argument in favour of public hearings for scientists who want to appeal the conclusions of the fraud office.

The interminable course of OSI's most public cases, coupled with the growing perception that the system is not fair, are enough to make the case that change is badly needed. In the past, the NIH have seen suggestions to remove OSI as a threat. They should see it as an opportunity to get out of an untenable situation in which the very agency that supports biomedical research also investigates it and, acting as judge and jury, is the arbiter of innocence or guilt.

Barbara J. Culliton

Optics beyond the limits

J. R. Heflin and A. F. Garito

BUCKMINSTERFULLERENE (C_{60}) has a wide variety of remarkable properties — for example, doped with alkali metals it becomes a superconductor second only to the high-temperature oxide superconductors, and under rapid compression it quickly converts into diamond. On page 225 of this issue¹, Tutt and Kost show how its unique optical properties allow C_{60} to be used as an optical limiter, transparent to low-intensity light, but nearly opaque above a critical intensity. Organic optical limiters such as C_{60} have potential uses as thresholding elements in optical digital processors and for protecting optical sensors from intense light.

Fullerene gains this capacity from its nonlinear optical properties: its response to light is not simply linearly dependent on the light's intensity. Over the past 10 years, there has been growing interest in the large, ultrafast nonlinear optical properties of conjugated organic and polymeric materials, associated with their delocalized π -electron orbitals²⁻⁷. Nonlinear optical properties are the basis of newly emerging photonics technologies in which light — streams of photons — works with, or even replaces, electrons in applications traditionally carried out by microelectronics. Organic materials are also advantageous in photonics because organic molecular design and synthesis is flexible, and crystalline and thin-film phases are relatively easy to prepare.

Nonlinear optical responses can in general be classified as resonant or non-resonant depending upon how close the optical frequencies employed are to the natural absorption frequencies of the material. Resonant and nonresonant nonlinear optical processes each have a characteristic set of applications for which they are best suited. Nonresonant responses are generally orders of magnitude faster and avoid attenuation of the optical signal, but they are also in general several orders of magnitude weaker than resonant responses. Therefore, resonant responses are preferred when the slower speed and increased optical loss are not detrimental to the application.

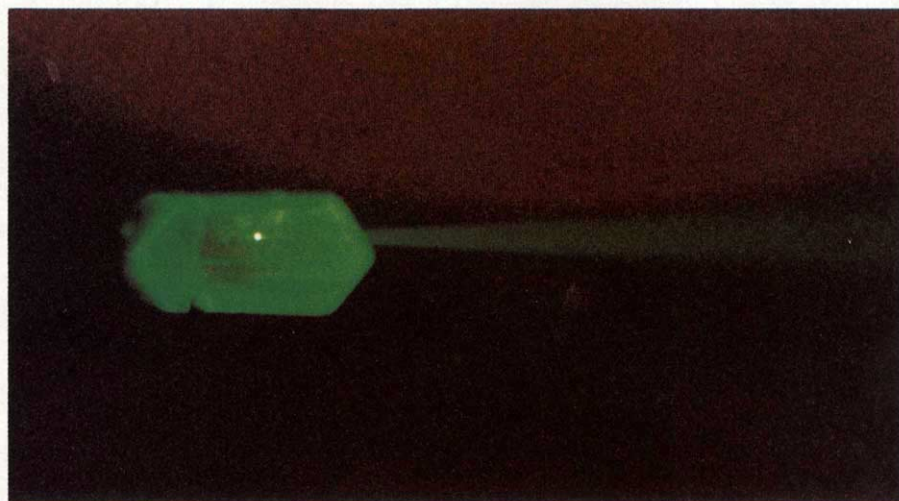
Nonresonant second-order processes, in which the response is proportional to the square of the applied electric fields, arise only in materials that lack a centre of symmetry. Second-harmonic generation, in which the material generates light at twice the frequency of the incident light (see figure), and the electro-optic effect, where an applied electric field changes the refractive index of the material and therefore alters the propagation of incident light, are two principal

examples of second-order processes. Conjugated organic rings and linear chains with diametrically opposed electron-donor and electron-acceptor groups have particularly large second-order responses².

In addition to preparation as single crystals or Langmuir-Blodgett films, great strides are being made³⁻⁷ in condensing these molecules into noncentrosymmetric assemblies as 'poled polymer films'. The active nonlinear optical molecules are dispersed in these glassy poly-

cesses in key organic and polymeric structures⁷. The insight gained from the correlation description is currently being applied to the design of new molecular structures and exploration of new enhancement mechanisms to achieve unusually large third-order optical properties which, in turn, are of keen interest for all-optical switches and modulators, especially in fibre optic networks^{3,7}.

Resonant third-order nonlinear optical processes in organic materials have been investigated for quite some time, for example, in studying saturable absorption and in the course of developing Q-switching laser dyes which produce short optical pulses by reducing the loss in a laser cavity at high intensity⁸. Satur-



Visible improvement: demonstration of the second-harmonic generation possible with the organic crystal 2-methyl-4-nitroaniline. The crystal is irradiated from the left by invisible infrared light (wavelength, 1,064 nm) from a laser and produces green light (532 nm).

mers and aligned by an electric field while at a high temperature. The ordering is retained after cooling, thus producing materials with large second-harmonic-generation and electro-optic coefficients in a relatively easy manner. These materials have been used in electro-optical modulators that can operate at frequencies of tens of gigahertz and require only a few volts for full modulation^{3,4,6,7}.

Nonresonant third-order processes, cubic in applied electric fields, occur naturally in all materials. A comparable understanding of the fundamental origin of the nonresonant third-order nonlinear optical response of conjugated organic materials has been achieved only recently. Theory³⁻⁵ shows that the strength of the effect is strongly affected by electron correlation due to electron-electron repulsion. It arises primarily from a previously undiscovered, strongly correlated, energetically high-lying two-photon state that has a large excited-state absorption cross-section. These calculations have been quantitatively verified through dispersion measurements of several third-order nonlinear optical pro-

cesses. The insight gained from the correlation description is currently being applied to the design of new molecular structures and exploration of new enhancement mechanisms to achieve unusually large third-order optical properties which, in turn, are of keen interest for all-optical switches and modulators, especially in fibre optic networks^{3,7}.

able absorption is the nonlinear transmission phenomenon complementary to optical limiting in which a decrease in the light absorption of a material is observed as the light intensity is increased. When a material is irradiated by intense light that is resonant with a strong optical transition, a large fraction of molecules will be excited into their upper electronic state. As long as there is not another strong transition at the same frequency starting from the excited state, the material can absorb no more light and so becomes transparent.

Organic structures can also exhibit optical limiting behaviour which requires special conditions that are the reverse of those for saturable absorption. For optical limiting to occur at a particular laser wavelength, the material must possess a weak but nonzero absorption at that wavelength. However, at the given wavelength, there must be a higher-energy transition with a large excited-state absorption cross-section from the state that is populated by the residual ground-state absorption. Thus, a single material can act as a saturable absorber for wavelengths of light within a strong

Shosaku Numa (1929–1992)

absorption band and as an optical limiter for wavelengths that are in a relatively transparent region.

Tutt and Kost¹ recognized that the large triplet excited-state absorption of C_{60} (ref. 9) makes it a good candidate for optical limiting. Each carbon atom in C_{60} is bound to its three nearest-neighbour carbon atoms with sp^2 hybrid σ -bonds. The remaining valence electron of each carbon contributes to a π -electron distribution that is delocalized over the entire sphere. The π -electrons are highly mobile and are responsible for both the large oscillator strengths of electronic transitions in conjugated structures and their large nonlinear optical responses. The relatively large size of the C_{60} sphere implies that the optical transitions both from the ground state and from the excited states should be quite strong. The threshold for optical limiting is dependent on the difference between the excited-state and ground-state absorption cross-sections. Indeed, Tutt and Kost find that, for green light with a wavelength of 532 nm where the excited-state absorption cross-section is significantly larger than the ground-state cross-section, C_{60} has an exceptionally low optical limiting threshold compared with many other materials known to be optically limiting at 532 nm.

The optical limiting threshold, loosely defined as the intensity at which the material becomes opaque, is lowest for large excited-state absorption cross-sections and long excited-state lifetimes. Presumably, the optical limiting threshold of C_{60} is so low because the lifetime of the triplet excited state is relatively long, tens of microseconds⁹ (10^{-6} s), as compared to singlet-state lifetimes on the scale of nanoseconds (10^{-9} s) and picoseconds (10^{-12} s). Furthermore, the excited-state absorption of C_{60} is more than three times larger at 400 nm than at 532 nm (ref. 9). One would therefore expect C_{60} to serve as an even more effective optical limiter at this wavelength. □

J. R. Heflin and A. F. Garito are in the Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

SHOSAKU Numa, the most prominent figure in the study of the molecular structure of receptors and ion channels in excitable membranes, died from cancer on 15 February 1992, shortly after his sixty-third birthday. His death at the height of his scientific productivity has come as a shock to cellular and molecular neurobiologists in particular, for Numa's work has transformed neurobiology over the past decade.

Born in Wakayama, Japan, Numa studied medicine at Kyoto University,



Numa: committed to research.

where he qualified in 1953, and he completed his training at Harvard Medical School. The next ten years were devoted to examination of the pathways that regulate the biosynthesis of fatty acids. While working at the Max Planck Institute for Cell Biology in Munich with Feodor Lynen, he discovered the regulation of fatty acid synthesis by acetyl-CoA carboxylase. Later he purified this enzyme and studied its allosteric regulation. His experience in Lynen's laboratory confirmed his commitment to research rather than medicine.

On returning to Kyoto in 1968, Numa established a vanguard group working on the molecular characterization of the proteins that mediate communication between neurons. Initially he concentrated on peptides that act as mediators of neural signalling. His group determined the gene structure of precursor molecules from which neurally active peptide hormones (such as corticotropin-releasing factor and opioid peptides) are derived, work which was in itself a landmark.

But Numa's greatest contribution was to the study of the molecular structure of membrane proteins — receptors, ion channels and ion pumps — that endow nerve and muscle cells with signalling properties. The early work centred on the receptor for nicotinic acetylcholine, which is a key element in synaptic communication between nerve and muscle cells. It was the first receptor-gated ion channel for which the complete primary structure was described, and its charac-

terization was the beginning of a body of research that has pointed to a common ancestral gene for the genes encoding the various transmitter-gated channels: the techniques involved in isolating complementary DNAs for this receptor were prototypical for studies elucidating the structure of a whole family of ligand channels.

Numa's group subsequently cloned and sequenced the gene for the sodium channel, a member of the other main class of ion channels, voltage-gated channels, which mediate communication within neurons. This development marked the beginning of the realization that voltage-gated channels also form a family with common structural building principles. Also to fall to Numa's insight and energy were the molecular basis of the functional heterogeneity of components mediating slow signal transduction, such as the muscarinic receptors and GTP-binding proteins, and the primary structures of ion pumps (Na^+/K^+ ATPase); this is another major class of membrane proteins required for electrical and chemical excitability.

Numa's research also took in the establishment of structure–function relationships. Collaborative efforts with membrane physiologists, combining site-directed mutagenesis, heterologous expression and biophysical analysis of channel function, identified, for example, the molecular difference between developmentally regulated acetylcholine receptor-channel subtypes and determinants of ion transport through this channel. With Walter Stühmer's group, identification of a putative voltage sensor and of the determinants of ion selectivity was achieved; with Kurt Beam's group, the molecular components of electromechanical coupling were delineated; and with Erwin Neher, the molecular basis of the heterogeneity of muscarinic acetylcholine receptor subtypes was discovered, and the functional expression of the calcium release channel demonstrated.

Shosaku Numa was a fascinating man. His high- and single-minded purpose, his striving to be accurate in every detail, and his seemingly limitless energy were all reflected in the quality of his scientific achievements and the clarity of his numerous plenary and award lectures. He leaves us with the framework for solving the three-dimensional structure and structure–function relations, as well as the evolutionary origins, of the membrane proteins that enable cells to communicate.

Bert Sakmann

Bert Sakmann is in the Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, D-6900 Heidelberg 1, Germany.

1. Tutt, L. W. & Kost, A. *Nature* **356**, 225–226 (1992).
2. Williams, D. J. (ed.) *Nonlinear Optical Properties of Organic and Polymeric Materials* (American Chemical Society, Washington DC, 1983).
3. Messier, J. et al. (eds) *Nonlinear Optical Effects in Organic Polymers* (Kluwer Academic, Boston, 1989).
4. Hann, R. A. & Bloor, D. (eds) *Organic Materials for Nonlinear Optics* (Royal Society of Chemistry, 1989).
5. Bowden, C. M. & Haus, J. W. (eds) *Nonlinear Optical Properties of Materials J. Opt. Soc. Am.* **B6**, 562–583 (1989).
6. Chiang, L. Y., Chaikin, P. M. & Cowan, D. (eds) *Advanced Organic Solid State Materials* (Materials Research Society, Pittsburgh, 1990).
7. Singer, K. D. (ed.) *Nonlinear Optical Properties of Organic Materials IV Proc. SPIE* **1560** (1991).
8. Sorokin, P. P., Luzzi, J. J., Lankar, J. R. & Pettit, G. D. *IBM J. Res. Dev.* **8**, 182–184 (1964).
9. Arbogast, J. W. et al. *J. phys. Chem.* **95**, 11–12 (1991).

Up the kinetic pathway

T. E. Creighton

THE question of whether protein folding is under kinetic or thermodynamic control takes a twist with the appearance on page 263 of this issue¹ of a paper by Baker, Sohl and Agard. Their report provides an example of a substantial kinetic block to folding in the case of α -lytic protease. It follows close on the demonstration by Carrell *et al.*² and Mottonen *et al.*³ of instances of the generation by folding of metastable protein conformations, in preference to the more stable forms to be expected from thermodynamic considerations.

It was once an article of faith of those working on protein folding that kinetic pathways must be involved; this view stemmed from the argument of the late Cyrus Levinthal⁴ that there are too many conformations of an unfolded protein for the native conformation to be adopted by random processes. Yet little hard evidence to support this belief has been found in the intervening 22 years, and the faith of some began to waver^{5,6}.

Proteins are certainly observed to re-fold nonrandomly⁵, but if the nonrandom pathway is crucial one should be able to block folding by interfering with that pathway. Most examples of proteins that do not fold are due to the competing process of aggregation of the unfolded protein, rather than a kinetic block. Whenever such a block has been introduced or identified, proteins have been found to use an alternative pathway with an overall transition state of just slightly higher free energy (an increase in free energy of the transition state of only 1.4 kcal mol⁻¹ decreases the rate of a reaction at room temperature by a factor of ten). A second prediction of kinetic pathways being important is that a protein might fold for kinetic reasons to a structure that is not the most stable thermodynamically (that is, to a metastable state). Yet whenever a folded protein is observed to be metastable, the more stable state is usually an insoluble precipitate, not the well-defined structure hoped for. So it is that the results described by Baker and colleagues, and by Carrell *et al.* and Mottonen *et al.*, come as a reward for the faith of those believing in the importance of kinetic pathways.

Bacterial proteases

A number of bacterial proteases from different families are synthesized as inactive precursors with N-terminal pro-segments that are necessary for biosynthesis of the active enzymes; the pro-segments are subsequently removed proteolytically to generate the active, native

protease. In the case of the subtilisin⁷ and α -lytic proteases⁸, the pro-regions consist of 77 and 166 residues, respectively. These pro-proteases unfold and refold *in vitro*, whereas the mature forms do not refold, even though unfolded α -lytic protease retains its three native disulphide bonds. The inability of proteases to refold is not new or surprising, for the first few molecules to regain activity tend to degrade the other molecules still unfolded. The mature forms of the bacterial proteases were shown to refold in the presence of added pro-segment^{9,10}, but this could be explained by the added pro-segments being inhibitors of the proteolytic activity¹¹. Another possible explanation was simply that the unfolded protein aggregated irreversibly, and the pro-segment might be playing the role of a chaperone¹².

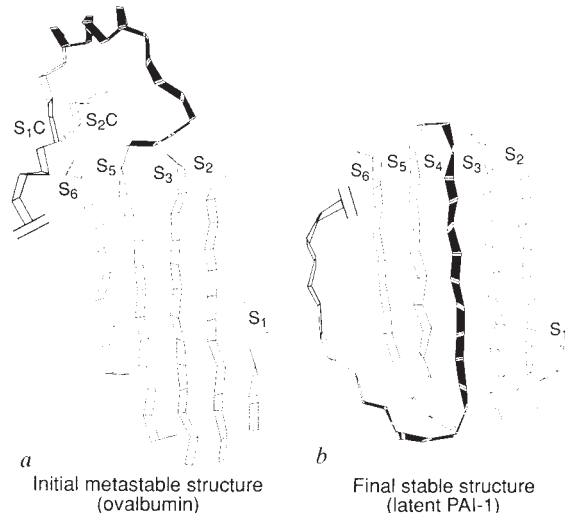
Baker *et al.*¹ now show these explanations to be incorrect in the case of α -lytic protease. Rather than being proteolyzed or aggregating, unfolded mature α -lytic protease under refolding conditions remains competent for refolding over a matter of weeks, but folds only when the pro-segment is added. Under refolding conditions, the unfolded mature α -lytic protease becomes somewhat compact and rather like the 'molten globule' state of proteins (probably best described as a collapsed form of the unfolded state). The negligible rate of refolding of mature α -lytic protease, under conditions where the folded protein is very stable, indicates that the free-energy barrier to its refolding is greater than 27 kcal mol⁻¹. The pro-segment must interact in some way with the unfolded state to lower the free energy of the overall transition state for folding by at least 9 kcal mol⁻¹ and to increase the rate of folding by at least 10⁷ times¹. This is the folding that normally occurs in the pro-protein in an intramolecular process, and the kinetic block was uncovered only by dissecting the protein into its two parts.

These observations demonstrate that the possible existence of a stable folded

protein conformation does not guarantee that the conformation will be kinetically accessible. The existence of such a kinetic block in this case implies that it could also exist with other proteins but not be apparent, perhaps because the proteins need to be dissected or because alternative, less stable conformations are generated by the kinetic process of folding.

Protease inhibitors

The other evidence for kinetic factors determining protein structure comes from work on the family of protease



a, The initial metastable structure of serpins generated by the folding process and, b, the final, more stable form that occurs after a slow conformational change (modelled on the crystal structures of ovalbumin¹⁴ and of latent plasminogen activator inhibitor-1 (PAI-1) (ref. 3), respectively). The main β -sheet A is illustrated, with the individual β -strands labelled S₁ to S₆; S₁C and S₂C are two strands of β -sheet C. The slow conformational transition from the metastable to the more stable form involves insertion of the darkened segment as the S₄ strand into β -sheet A, which in the case of PAI-1 requires dissociation of strand S₁C from β -sheet C. The darkened segment is the active-site region of serpins that function as protease inhibitors. The active inhibitor conformation is believed to be an early intermediate stage along the transition to the stable form, in which the darkened segment has a conformation complementary to that of the protease active site². The polypeptide chain tends to be cleaved by the protease, when the liberated segment with the new α -carboxyl group inserts as strand S₄ into sheet A, without perturbing the other segment or strand S₁C (ref. 13). (Figure kindly provided by Robin Carrell.)

inhibitors known as the serpins, which includes the physiologically important α_1 -antitrypsin (α_1 AT), antithrombin and plasminogen activator inhibitor-1 (PAI-1). These inhibitors function by insertion of particularly mobile segments of their polypeptide chain into the active sites of their target proteases², where they are susceptible to cleavage. Cleavage causes conformational changes, and the cleaved forms are less susceptible to irreversible denaturation. Substantial conformational changes in other proteins have always been found primarily to involve changes in the orientations of relatively rigid subunits or domains, but the crystal

structure of cleaved α_1 AT showed the amino and carboxyl groups originally linked by the cleaved peptide bond to be at opposite ends of the single domain, some 70 Å apart¹³. The segment of polypeptide with the new carboxyl group was hypothesized to have changed its conformation dramatically after the cleavage, and to have become incorporated as a sixth strand in the centre of a previously five-stranded β -sheet, one of three β -sheets in the structure and designated A (ref. 13). That such a conformational change took place was confirmed by the crystal structure of uncleaved ovalbumin¹⁴, which is a member of the serpin family, although not a protease inhibitor; the corresponding β -sheet of ovalbumin contains only five β -strands (see figure). Also, the more stable, presumably six-stranded, β -sheet form of serpins could be mimicked by adding to the uncleaved forms an extraneous peptide corresponding to the sixth strand⁵.

These observations had dramatic implications for the flexibilities of protein structures and of β -sheets. But they were not of immediate significance for protein folding, because the conformational changes took place only after cleavage of the polypeptide chain. Similar changes had been known for some time to occur in human PAI-1, however, without cleavage of the polypeptide chain. PAI-1 is synthesized in a form that is active as an inhibitor, but relatively unstable. The active form slowly converts, with a half-time of about 1 hour at 37 °C, to a form that is more stable, but inactive as an inhibitor, and is known as the latent form¹⁵. If this latent form is unfolded in denaturants¹⁶ or at high temperatures¹⁷ and then refolded, the active metastable form is regenerated. Carrell *et al.*² showed that similar transitions from active forms to more stable forms could take place in other intact serpins in low

concentrations of denaturant, and hypothesized that this transition involves the insertion of the intact polypeptide chain as a sixth strand into the β -sheet as far as the length of the flexible portion would permit. The crystal structure of the latent form of PAI-1, described by Mottonen *et al.*³, shows the sixth strand to be fully inserted into the β -sheet (see figure). Surprisingly, the extra length of flexible polypeptide chain needed for complete insertion of the β -strand was obtained by removing an edge β -strand from another, four-stranded β -sheet, designated C.

The structural acrobatics of these proteins are unprecedented, but the observations are even more notable for demonstrating that the folding process of PAI-1, both *in vivo* and *in vitro*, produces a metastable structure, probably containing one five-stranded β -sheet (A) and another four-stranded sheet (C). This folded conformation is not very

stable, and it converts to a more stable form with one six-stranded and one three-stranded β -sheet (see figure).

It is probably pertinent that the examples reported here are of proteases and protease inhibitors. The kinetic block in folding of the mature bacterial proteases may well prevent premature formation of a destructive protease, and the unusual mobility of the conformations of the serpins is undoubtedly important for their inhibitory properties². The interactions between proteases and their inhibitors are probably also why both provide the most dramatic evidence for positive selection, in increasing their rates of evolutionary divergence¹⁸. It will be surprising if study of them is not rewarded with evidence for other phenomena not apparent in other classes of proteins. □

T. E. Creighton is in the European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, Germany.

PHYCOLOGY

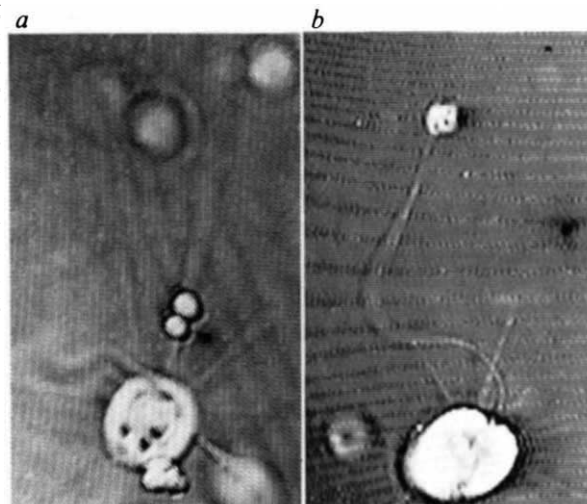
What the haptonema is for

Ralph A. Lewin

M. KAWACHI and colleagues (*Phycologia* **30**, 563–573; 1991) have neatly solved (or partly solved) a problem that has worried phycologists for years: why certain chrysophytes, sometimes called haptophytes, or 'grappling algae', have a haptonema.

This organelle has been known for more than a century, since it is just visible by light microscopy. It is a filament something like a flagellum extending between the paired cilia of certain unicellular algae, but is clearly not involved in their swimming mechanism. Early in the history of electron microscopy its axis was shown to be made up of microtubules. But whereas flagellar axonemes always have nine pairs of microtubules, usually surrounding a central pair, the haptonema has six or seven enclosed in two or three sheathing membranes. Short haptonemata are little more than spurs; long ones may exceed the cell in length, and are normally coiled. People had postulated that a haptonema might be involved in feeding, but many of us discounted this idea because haptophytes are generally photosynthetic. Little did we suspect what dark deeds could be done by such little plants.

Kawachi *et al.* show how one prymnesiophyte (the class name now preferred) uses the haptonema to catch and eat bacteria and other little morsels. The species they looked at was *Chrysochromulina hirta*. As the cell swims, the organelle is extended straight ahead. Small particles stick to it, and are caused to slither down its surface towards a site some micrometres from the base. After a few have accumulated in a kind of bacterial bolus, apparently held together by slime, the lump is sent back up to the



Chrysochromulina hirta fishing with the haptonema (in this instance for latex microspheres 0.88 micrometres in diameter). *a*, Accumulation towards the base of the haptonema of the latex 'bolus'; *b*, the spheres, here appearing as one, at the tip of the filament, poised for their journey to a buccal area on the side of the cell. (Photographs courtesy M. Kawachi.)

1. Baker, D., Sohl, J. L. & Agard, D. A. *Nature* **356**, 263–265 (1992).
2. Carrell, R. W., Evans, D. L. & Stein, P. E. *Nature* **353**, 576–578 (1991).
3. Mottonen, J. *et al.* *Nature* **355**, 270–273 (1992).
4. Levinthal, C. *J. chim. Phys.* **65**, 44–45 (1968).
5. Kim, P. S. & Baldwin, R. L. A. *Rev. Biochem.* **59**, 631–660 (1990).
6. Harrison, S. C. & Durbin, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4028–4030 (1985).
7. Ikemura, H., Takagi, H. & Inouye, M. *J. biol. Chem.* **262**, 7859–7864 (1987).
8. Silen, J. L., Frank, D., Fujishige, A., Bone, R. & Agard, D. A. *J. Bact.* **171**, 1320–1325 (1989).
9. Zhu, X. *et al.* *Nature* **339**, 483–484 (1989).
10. Silen, J. L. & Agard, D. A. *Nature* **341**, 462–464 (1989).
11. Baker, D., Silen, J. L. & Agard, D. A. *Proteins* **12**, 339–344 (1992).
12. Winther, J. R. & Sørensen, P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9330–9334 (1991).
13. Loeberman, H., Tukuoka, R., Deisenhofer, J. & Huber, R. *J. molec. Biol.* **177**, 531–556 (1984).
14. Stein, P. *et al.* *Nature* **347**, 99–102 (1990).
15. Franke, A. E. *et al.* *Biochim. biophys. Acta* **1037**, 16–23 (1990).
16. Hekman, C. M. & Loskutoff, D. J. *J. biol. Chem.* **260**, 11581–11587 (1985).
17. Katagiri, K. *et al.* *Eur. J. Biochem.* **176**, 81–87 (1988).
18. Creighton, T. E. & Darby, N. J. *Trends biochem. Sci.* **14**, 319–324 (1989).

haptonemal tip (see figure). Seconds later, the whole filament curves, bending back on itself, and conveys its meatball to a buccal area on the side of the cell, where a bald patch appears in the scaly cell coat. There it is ingested into the cell. The haptonema then straightens out again and recommences fishing.

The mechanisms underlying the migration of the attached particles are still unknown, although they may parallel a similar phenomenon that occurs on the surfaces of algal flagella. Unknown, too, is the means by which the haptonema neatly conveys its mouthful to the right site (perhaps ATPase 'arms' on the sides of the microtubules are involved, as they are in flagellar movement).

Algae can no longer be regarded simply as primary producers, because even among the apparently vegetative components of phytoplankton pastures there are many facultative predators. Some, such as the colourless dinoflagellates (for instance the giant cells of *Noctiluca*), are

obligate predators, but in the past few years we have learned that many smaller dinoflagellates, evidently with good functional photosynthetic apparatus, can also use solid food by extracellular or intracellular digestion of prey cells; a recent example is *Amphidinium cryophilum*, as documented by L. Wilcox and G. Wedemeyer in *Journal of Phycology* (27, 600–609; 1991).

Among predatory algae we now have to include prymnesiophytes such as *Chrysochromulina* which are as small as 5 micrometres in diameter. But what happens in nature? Could this sort of feeding go on all the time? Or does it merely constitute a source of midnight snacks or supplementary food, tiding the algae over periods when photosynthesis is precluded at night, or when the cells sink below the euphotic zone? □

Ralph A. Lewin is in the Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, California 92093, USA.

CELL BIOLOGY

A family of fusion proteins

John Aitken

ON page 248 of this issue¹, Carl Blobel and colleagues describe results that will attract the attention of any cell biologist interested in the fundamental process of membrane fusion. Their data suggest that fusion events as disparate as the fertilization of the mammalian ovum, and the penetration of an enveloped virus into its target cell, may involve highly conserved motifs in one particular family of membrane receptors.

The general aim of this work is to understand the molecular basis of the event with which we all commence our voyage through life — the fusion of the spermatozoon with the vitelline membrane of the oocyte. This is a highly cell- and species-specific process, involving a precisely defined region of the sperm plasma membrane that acquires the capacity to recognize, and fuse with, the vitelline membrane of the egg. However, it is not until the spermatozoon has been stimulated to undergo a secretory event known as the acrosome reaction, following contact with a glycoprotein (ZP3) in the zona pellucida, that it suddenly acquires this capacity.

Using the guinea-pig as an animal model, Paul Primakoff, Diana Myles and colleagues have pioneered the analysis of the molecular basis of this membrane fusion event. Their starting point was a monoclonal antibody (PH-30) that disrupts the fusion of acrosome-reacted guinea-pig spermatozoa with the vitelline membrane of the oocyte². The antibody precipitates two tightly coupled, im-

munologically distinct subunits (α and β) that behave as a single integral membrane protein. The complex is expressed on the posterior surface of the mature sperm head, in the appropriate location for molecules involved in oocyte fusion. Moreover, both the α and the β subunits are generated as large precursor proteins that seem to be modified by proteolytic cleavage during testicular differentiation and subsequent epididymal maturation of the spermatozoa³. Because the completion of PH-30 processing is achieved as spermatozoa acquire the competence to fertilize the egg, there had been an attempt to find a causal link between the two events³.

The interesting notion that the PH-30 complex might share similarities with viral fusion proteins has previously been proposed, the rationale being that both are integral membrane structures consisting of tightly coupled subunits, generated as large precursor proteins³. Blobel *et al.*¹ now provide sequence data to support this hypothesis.

The α subunit is composed of a peptide core of 289 amino acids, with a single membrane-spanning domain towards the C terminus, a large extracellular domain and a short cytoplasmic tail. A putative viral fusion peptide has been identified on the α subunit, and it appears to be somewhat similar in sequence to a potential fusion peptide on the rubella virus. A feature of the fusion protein motif on PH-30 α is its relative hydrophobicity which, in conjunction

with the transmembrane anchoring segment on the same subunit, should permit the molecule to interact simultaneously with the plasma membranes of both the spermatozoon and the oocyte, thereby promoting fusion. Other features of the PH-30 α fusion peptide that are shared by viral fusion proteins include central proline residues and the fact that this stretch of amino acids can be modelled as a 'sided' α helix with most of the bulky hydrophobic groups on one face.

The sequence data on the β subunit indicate that it contains a 353-amino-acid peptide core, with two potential sites for N-linked glycosylation and a single membrane-spanning domain¹. But the protein's most notable feature is the presence of an integrin-binding 'disintegrin' domain (the disintegrins are a large family of highly conserved platelet-aggregation inhibitors found in snake venom). In the context of fertilization, the significance of this similarity stems from studies suggesting that the egg plasma membrane contains an integrin-like receptor⁴ which could interact with disintegrin domains on the β subunit of PH-30. Consistent with this hypothesis, micromolar concentrations of the integrin-binding peptide, RGD, can inhibit sperm-oocyte recognition and fusion in mammals^{4,5}. Even though the disintegrin motif of PH-30 β does not contain an RGD sequence, there are many examples of integrin ligands in which this particular tripeptide is absent.

So the report of Blobel *et al.* points to the existence of a unique membrane recognition/fusion complex in mammalian spermatozoa, one which is centrally involved in sperm-oocyte fusion. Intriguingly, the PH-30 complex appears to consist of two subunits, stemming from a common ancestral protein, in which the tasks of cell recognition and membrane fusion have become dissociated¹. The β subunit seems to have assumed the key function of cell recognition, using a disintegrin-like motif to target integrin receptors on the egg plasma membrane; conversely, the α subunit is specialized for membrane fusion and contains a fusion peptide with many structural similarities to viral fusion proteins. Thus it seems that viral infection and fertilization, nature's own masterpieces of gene-transfer technology, use similar, highly

1. Blobel, C. P., Wolfsberg, T. G., Turck, C. W., Myles, D. G., Primakoff, P. & White, J. M. *Nature* **356**, 248–252 (1992).
2. Primakoff, P., Hyatt, H. & Tredick-Kline, J. J. *Cell Biol.* **104**, 141–149 (1987).
3. Blobel, C. P., Myles, D. G., Primakoff, P. & White, J. M. *J. Cell Biol.* **111**, 69–78 (1990).
4. Bronson, R. A. & Fusi, F. *Fertil. Steril.* **54**, 527–529 (1990).
5. Bronson, R. A. & Fusi, F. *Biol. Reprod.* **43**, 1019–1025 (1990).
6. Aitken, R. J., Ross, A., Hargreave, T., Richardson, D. & Best, F. *J. Androl.* **5**, 321–329 (1984).
7. Aitken, R. J., Irvine, D. S. & Wu, F. C. *Am. J. Obstet. Gynecol.* **164**, 542–551 (1991).

conserved mechanisms.

The new data also carry implications for the diagnosis and treatment of male infertility. Defective sperm function is the most frequent single cause of human infertility which, for want of basic information on the precise molecular basis for the loss of fertilizing potential, is currently untreatable. But studies employing an *in vitro* fertilization assay, in which human spermatozoa are fused with hamster oocytes, have demonstrated that a large proportion of such cases involve the failure of acrosome-reacted spermatozoa to fuse with the vitelline membrane of the oocyte^{6,7}. The results of Blobel and colleagues take us SUPRAMOLECULAR CHEMISTRY

one step closer to a molecular definition of male infertility, because with their sequence data, the search can now commence for analogous fusion proteins in the plasma membranes of normal and defective human spermatozoa. Should such proteins be found, their key involvement in fertilization would also make them prime candidates for the development of fertility regulating vaccines, which could find application in both men and women. □

John Aitken is in the MRC Reproductive Biology Unit, Centre for Reproductive Biology, 37 Chalmers Street, Edinburgh EH3 9EW, UK.

New synthetic receptors

Leonard F. Lindoy

THE revolution in the development of synthetic receptor molecules recognized by the award of the 1987 Nobel prize in chemistry to Charles Pedersen, Donald Cram and Jean-Marie Lehn continues. Three reports in *Angewandte Chemie (Int. edn Engl.)* give details of new receptors (hosts) incorporating boron atoms in their structures. All are cyclic compounds containing a central binding cavity and in each case their synthesis represents an important and novel addition to the expanding field of synthetic host-guest chemistry with implications for the future development of more

efficient molecular catalysts. A common feature of each study is the use of molecular architecture to achieve spatial and/or electronic complementarity between the host and guest species. New receptors, adept at the selective capture of anions, are the result.

The first report by Hawthorne and colleagues (X. Yang, C. B. Knobler and M. F. Hawthorne *Angew. Chem.* **30**, 1507–1508; 1991) describes the synthesis of 12-mercurocarborand-4 (Fig. 1), the initial representative of a new class of rigid macrocycles that are capable of binding an anion such as chloride in the central cavity. The chloride-ion complex can be considered to be a charge-reversed analogue of the alkali-metal complexes of the well-studied cyclic polyether 12-crown-4. The very stable chloride complex of 12-mercurocarborand-4 is obtained in high yield simply by reacting the dilithiated borane cage 1,2- $C_2B_{10}H_{12}$ with mercuric chloride. The authors propose that the chloride ion serves as a template in cyclizing the mercury ions and their carborane ligands; an X-ray diffraction study shows that the cyclic tetrameric product contains the chloride in the central cavity with each mercury distorted towards it. This arrangement results in unprecedented square-planar coordination of the central chloride ion by the four mercury sites. It appears that the electron-withdrawing properties of the carborane cages greatly enhance the attraction of the mercury centres for the central chloride anion.

A key observation is that reaction of the chloride complex with silver ions results in removal of the chloride from the central cavity without causing the host molecule to decompose. The shape and electronic nature of the cavity appear suitable for the binding of a

range of other inorganic and organic anions; and it seems likely that the system might act as a homogeneous catalyst for promoting particular reactions, a possibility now being explored.

Reetz and coworkers describe in a pair of reports (M. T. Reetz, C. M. Niemeyer and K. Harms **30**, 1472–1474, 1474–1476; 1991) the synthesis of the first examples of a new class of selective boron-containing hosts which are capable of performing the dual role of binding both a cation and an anion *simultaneously* using different sites within the host. This has been achieved by incorporating both a cyclic crown ether backbone and a boron-containing (electron starved) site in the host (Fig. 2). Such a molecule may bind a cation using the polyether crown ring and either leave the corresponding counter anion unbound or bind it simultaneously at the boron site. The latter effect can be employed to achieve 'salt selectivity'. For example, this host shows remarkable selectivity in preferentially binding KF from a mixture of KF, KCl, KBr and KI. In this case the discrimination is associated with the simultaneous binding of cation and anion to the host, a situation shown to persist in the solid state by means of X-ray diffraction. In solution, the binding of the potassium and the fluoride to the host is synergistic: in part, the arrangement adopted in the KF adduct is undoubtedly influenced by the strong affinity of the fluoride ion for the boron centre, although a number of other factors are also important.

Reetz and coworkers also show that their host binds alcohols and amines simultaneously (as well as selectively). The overall reaction (see Fig. 3) involves the spontaneous deprotonation of the alcohol and corresponding protonation of the amine to produce alkoxy and alkyl ammonium ions; the former binds to the boron receptor site and the latter hydrogen bonds to the oxygens of the crown polyether ring (Fig. 4). Thus, in mimicry of many biological systems, the synthetic host is capable of recognizing and binding more than one organic guest molecule. The initial experiment in-

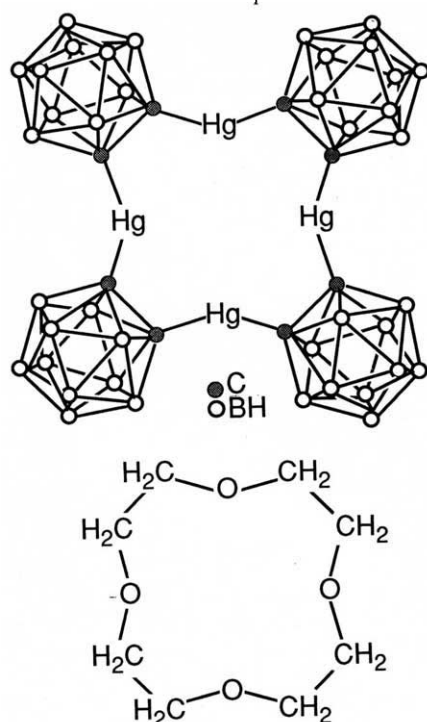


FIG. 1 12-mercurocarborand-4 (top) binds chloride anions in its interior just as its charge-reversed analogue 12-crown-4 (below) binds alkali metal atoms.

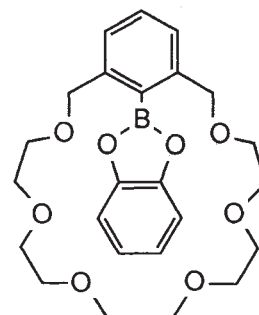


FIG. 2 This difunctional receptor is capable of simultaneous binding of a cation to its polyether (crown) ring and an anion to its boron site (see also Fig. 4).

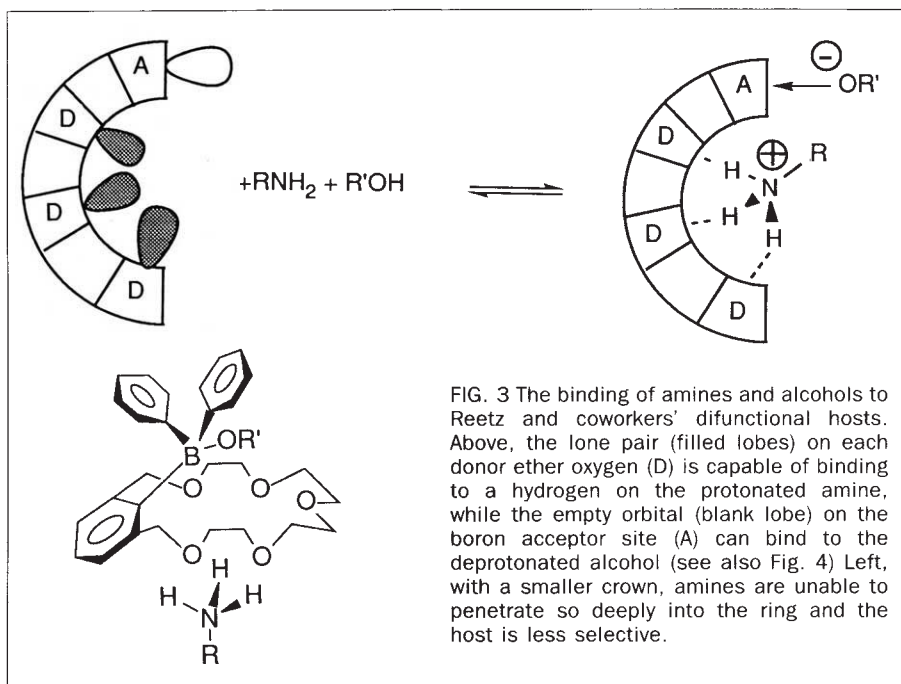


FIG. 3 The binding of amines and alcohols to Reetz and coworkers' difunctional hosts. Above, the lone pair (filled lobes) on each donor ether oxygen (D) is capable of binding to a hydrogen on the protonated amine, while the empty orbital (blank lobe) on the boron acceptor site (A) can bind to the deprotonated alcohol (see also Fig. 4). Left, with a smaller crown, amines are unable to penetrate so deeply into the ring and the host is less selective.

involved the interaction of the difunctional host with a 1:1 mixture of methanol and benzylamine in dichloromethane. An X-ray structure of the product shows that there is communication between the guests: the ammonium ion is contained in the cavity of the crown ring and is hydrogen bonded to the oxygens of the crown as well as to that of the methoxy group (which is bound at the boron acceptor site). Competition experiments involving benzylamine and two alcohols

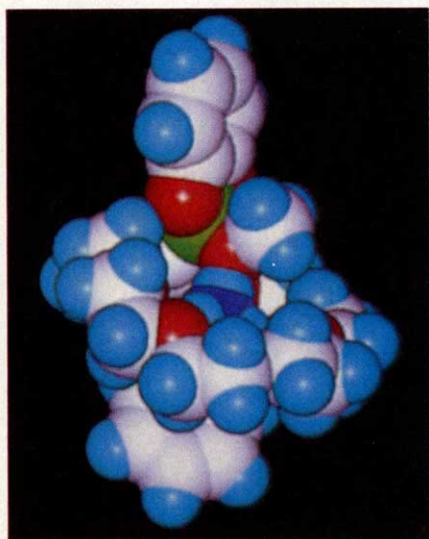


FIG. 4 Space-filling model of Reetz and coworkers' difunctional host bound to methanol and benzylamine. Carbon, white; hydrogen, pale blue; boron, green; oxygen, red; and nitrogen, purple. The methanol (in its methoxy form) is tucked into the fold to the right of the boron. The ammonium cation is tucked into the ether ring just below, its hydrogens bonded to the ether oxygens; the benzyl moiety extends below the ether ring. There is a further hydrogen bond between the ammonium ion and the methoxy oxygen.

indicate that selective alcohol binding occurs. For example, methanol binding is favoured over ethanol binding by a ratio of approximately 3:1. Such potentially useful molecular recognition appears to be largely controlled by steric factors.

Because the X-ray structure shows incorporation of benzylamine in the crown's cavity, the authors expected that the host might recognize amines selectively based on their size. Again, competition experiments with benzylamine and α -methylbenzylamine together with different alcohols show that, in each case, only a benzylamine/alcohol combination was selected.

Similar competition experiments involving a smaller version of the host (incorporating five oxygen atoms in the crown ring, instead of six) resulted in lower amine selectivity. Reetz and coworkers ascribe this to the inability of the respective ammonium groups to penetrate as deeply into this smaller crown (Fig. 3). This latter observation is in accordance with the X-ray structure of the ethanol/benzylamine adduct. But this smaller host yields increased selectivity for methanol; for example, in association with benzylamine, methanol is bound exclusively over ethanol.

The present innovations in receptor design have wider implications. These include the development of new anion sensing devices, new chromatography materials and new boron-containing reagents for use in neutron-capture therapy for the treatment of cancer. □

Leonard F. Lindoy is in the Department of Chemistry and Biochemistry, James Cook University, Townsville 4811, Queensland, Australia.

Cries to heaven

SOUND, says the physics syllabus, travels in straight lines. Daedalus disagrees. He points out that air molecules move in the Earth's gravitational field. They travel not in straight lines, but in parabolic paths like any other projectile. The path of a sound wave, being composed of innumerable molecular trajectories, must be parabolic too.

On this argument, you might expect sound to have the same ballistics as a gun whose shell travelled at the velocity of an air molecule. Its maximum range would be about 25 km, achieved by launching it at 45° to the horizontal. This isn't quite right. A shell slows up as it climbs, but sound doesn't — its speed depends purely on the local air conditions. This extends its range dramatically. High enough angles of launch should project it hundreds of kilometres.

Its maximum practical range, says Daedalus, is set by the height of the atmosphere. 400 km up, the mean free path is about 25 km. A sound wave which reached this height would launch a set of molecules on purely ballistic trajectories. They would curve over in a huge collisionless parabola until, falling back into the denser atmosphere below, they launched the sound on its return journey to Earth. This atmospheric 'mirror-ceiling' limits the horizontal range of sound to about 300 km. Gunfire has occasionally been heard over this distance.

Daedalus likes the idea of launching a sound into the sky, and having it boom down from the heavens in a narrow target zone many kilometres away. This novel form of directional communication would be welcomed by unscrupulous advertisers and evangelists seeking to demonstrate celestial approval of their claims. But the required phased-array transmitter, with its banks of high-powered loudspeakers distributed over many hectares, could be better employed in atmospheric research.

For this purpose the transmitter would launch an intense continuous tone. The tone would reach the distant receiving station with a complex and rapidly changing amplitude and phase, reflecting the many vicissitudes of its long stratospheric journey: wind shear, Doppler shifts, fluctuations of temperature and so on. The transmitter and receiver would be linked by at least two sonic paths, one a normal parabola and the other bounced off the atmospheric ceiling. The differences between them would be informative, especially over a range of frequencies. A huge volume of atmosphere could be neatly interrogated from the ground. Weather forecasting might be transformed. Nobody outside the system would be deafened. David Jones

Octopod 'ballooning' response

SIR — Very little is known about the deep-sea octopuses of the order Cirroctopoda¹. Observations of live animals were rare until recently because aquarium studies were possible only in the case of the flapjack devilfish *Opisthoteuthis*². Deep-sea cameras subsequently allowed cirrate octopods to be photographed in their natural habitat^{3,4}. We have now observed these animals

previously observed³. The submersible was brought sufficiently close to touch the animal with the grab arm. This contact triggered a spectacular reaction, which we describe as the 'ballooning' response — from the inverted umbrella attitude (*a* in the figure) the animal changed in about 15 seconds to a pump-kin shape. The process is summarized below, based on our video recording of the animal.

During hovering, the outer-arm web was not fully expanded so that the so-called intermediate web was visible as a vertical connection between the main (outer) sheet of the web and the curved muscular arm trunk (*a*). On contact, the animal reacted by changing the curvature of the arms and fully expanding the web so that the arm crown took on a bell shape. The arm trunks continued to curve inwards while the web sectors lying between the arms bulged out strongly; the intermediate-web bands became high dividing membranes above each arm trunk. The arm tips then disappeared beneath the web (7 seconds after the beginning of the response). The ballooning increased for another 10 seconds, the curved arms being brought together more closely. The arm tips finally reappeared due to a peristaltic wave generated by a coordinated arm flexion moving backwards from the arm tips (*b*). While the animal remained stationary, some water must have escaped at the surface of the web

smoothed; 30 seconds after the beginning of the response, the grooves between the web sectors disappeared and the body of the animal reappeared above the shrinking web sphere. After 43 seconds the fins, which had remained immobile during hovering and ballooning, were lifted, starting the backwards swimming movement: this completely 'emptied' the trailing arm crown.

We can only guess the function of the ballooning response. In an environment where downwelling light does not penetrate, tactile stimuli are likely to be important. The 'transformation' achieved by the ballooning response might have a stunning or disorientating effect on a potential predator in a first

(possibly accidental) contact with the cirrate. The ballooning web, by providing 'imprecise' tactile cues, could thus be a defence mechanism.

S. V. BOLETZKY

URA 117 CNRS,
Observatoire Océanologique de
Banyuls,
Laboratoire Arago,
66650 Banyuls-sur-Mer,
France

M. RIO

Centre Sciences de la Terre,
Université de Lyon I,
69622 Villeurbanne, France

M. ROUX

Laboratoire Sciences de la Terre,
Université de Reims,
BP 347 et URA
157 CNRS,
51062 Reims, France

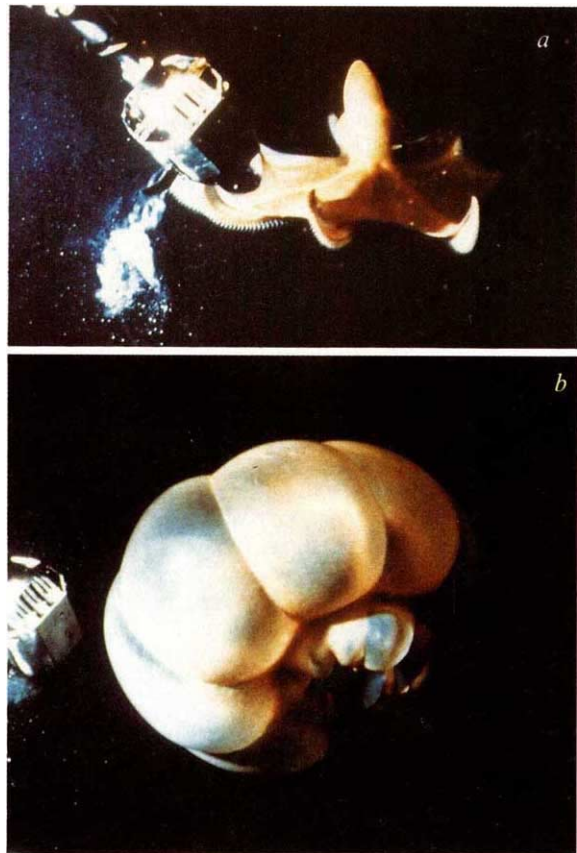
1. Young, J. Z. *Phil. Trans. R. Soc.* **B325**, 189–237 (1989).
2. Pereyra, W. T. *Pacif. Sci.* **19**, 427–441 (1965).
3. Roper, C. F. E. & Brundage, W. L. *Smithson. Contr. Zool.* **121**, 1–46 (1972).
4. Percy, W. G. & Beal, A. *Deep-Sea Res.* **20**, 107–108 (1973).
5. Roux, M. et al. *Bull. Soc. géol. France* **162**, 675–685 (1991).

Vortex air rings

SIR — Commenting on Aref and Zawadzki's simulation-based studies on the linking of vortex rings¹, J. D. S. Jones in *News and Views*² asked whether the vortex-linking could be done experimentally. I recently observed and photographed underwater 'air rings', which I believe to be an example of this linking.

In November 1991 I rode in the submarine *Atlantis*, a commercial sight-seeing submarine based at Kailua-Kona, Hawaii. A diver accompanied the submarine to a depth of about 50 feet, feeding fish, watching for sharks and blowing air rings. The diver was facing up when he produced each air ring, so each ring rose purely vertically. On one occasion (left-hand photo over page) the diver produced two air rings in quick succession, followed by a third. As he blew a fourth ring (right-hand photo) and destroyed it with his hand, the second ring caught up with the first and linked with it. The process was fascinating to watch: one portion of the second ring began to rise more quickly than the rest, as if the first ring were pulling it towards it. The corresponding portion of the first ring also began to rise more quickly than the rest of its ring, although not so markedly as in the second ring. Soon the second ring caught and linked to the first ring, forming a single large ring of irregular shape. After a few more seconds the situation became more complicated, with the large ring becoming unstable and the third ring catching up with the other two.

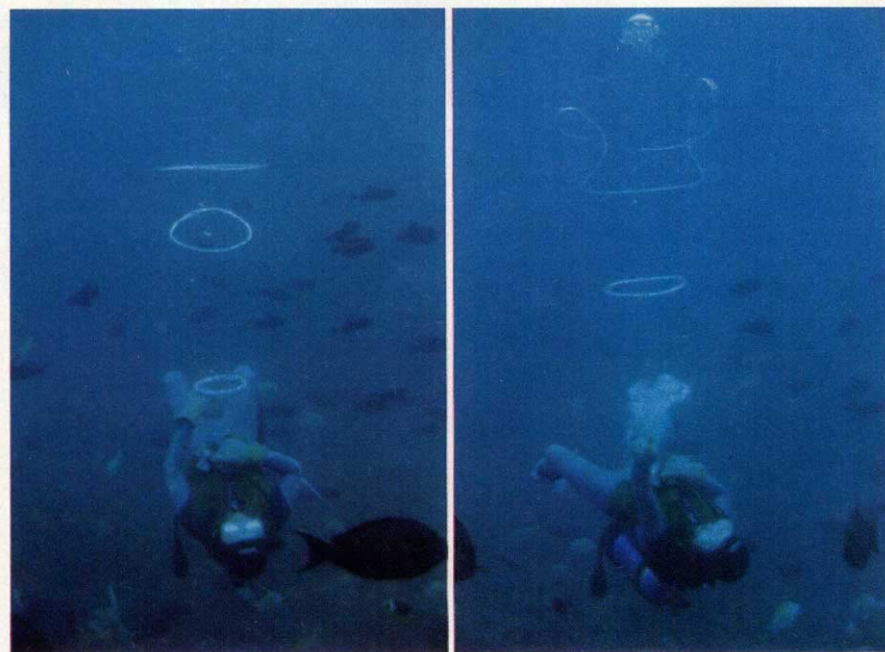
Although this technique seems to be



a, A cirrate octopod hovering in midwater in the inverted umbrella position; the grab arm of the submersible has not yet made contact with the animal. *b*, The ballooning response at its peak, the outer web being fully extended, with arm tips reappearing at lower right. Note the proximal parts of the muscular arm trunks which are bent to form arcs (visible through the extended membranes).

from a submersible, providing the possibility for interacting with individual cirrates. This avenue for investigation could be fruitful in investigating the behaviour of other free-ranging, deep-sea animals.

During the French CALSUB programme⁵ in February–March 1989 on board the research vessel *Le Suroît* in the southwestern Pacific, the crew (including M. Rio) of the diving saucer *Cyana* (SP 3000) saw a cirrate octopod (possibly of the genus *Cirrothauma*) off the northeast coast of Lifou Island (Loyalty Ridge) at a depth of 2,880 metres, just above the sea floor. The animal hovered in the horizontal 'inverted umbrella' position (see figure)



Underwater air rings. Left, three rings rising vertically; right, top rings become linked.

limited to studying vortices that travel in parallel planes, it does show that vortex linking is easy to demonstrate experimentally.

RONALD L. RIVEST
MIT Department of Electrical
Engineering and Computer Science,
Cambridge,
Massachusetts 02139, USA

AREF REPLIES — I had previously seen pictures of these 'air-rings' (usually called 'vortex ring bubbles' in the fluid mechanics literature) that divers can produce. There is a recent paper by Lundgren and Mansour³ where such bubbles are studied and these authors mention a photograph in *National Geographic Magazine*⁴. Apparently, whales and dolphins also blow the rings.

I am not sure that what Rivest saw is the same process of linking that we reported in our paper¹. In our study, the intermediate state of interest consists of two rings connected as successive links in a chain, whereas he reports the formation of 'a single large ring of irregular shape'. The linking that we were interested in is a variant of the more general phenomenon of reconnection, where two rings can, indeed, form a single, irregular ring for a while. I think what Rivest is seeing (see right-hand photo) is more closely related to the type of intermediate state that we have in our Fig. 1d–e rather than the linked state in our Fig. 4d. We actually believe that the ellipticity of the initial rings is essential for the process. In our simulations we did not succeed in producing linking starting from two circular rings.

I am not sure how one would blow elliptical vortex ring bubbles, but if it were possible, it would make a spectacu-

lar real-world analogue of the numerical solutions.

HASSAN AREF

Institute of Geophysics and
Planetary Physics,
University of California, San Diego,
La Jolla, California 92093, USA

1. Aref, H. & Zawadzki, I. *Nature* **354**, 50 (1991).
2. Jones, J. D. S. *Nature* **354**, 21 (1991).
3. Lundgren, T. S. & Mansour, N. N. *J. fluid Mech.* **224**, 177 (1991).
4. Earle, S. A. & Giddings, A. *Natn. geogr.* **149**, 578 (1976).

Neanderthal dates debated

SIR — We cannot agree with some of Stringer and Grün's speculations¹ stemming from the redating of the late Neanderthal from Saint-Césaire. We wonder why a thermoluminescence date of 36,300 years before present (BP), several thousand years older than previously thought, can justify the contention that "Neanderthals probably went out with a whimper, not a bang"¹.

The new date means that 12,500 years or more may separate Saint-Césaire from Cro-Magnon, the site of the earliest definitive 'modern' human crania from western Europe², if these burials are Upper Perigordian and not Aurignacian, as has been suggested³. How can this new, expanded timespan show an ancestor-descendant relationship to be chronologically untenable in western Europe? We believe that the western European dates may provide the best evidence for a region where there was a transition to modern populations with significant local genetic input from Neanderthals, in that there is now more than enough time for the process of

evolution due to natural selection to proceed at a reasonable rate. For instance, the rates of change between Neanderthals and early 'modern' Europeans for a truly diagnostic feature such as anterior tooth size are small relative to other recent evolutionary changes. These rates are one-half those between early and late Upper Palaeolithic samples and one-tenth the rate of change between Europeans of the Mesolithic and Neolithic who are only 3,500 years apart (D.W.F., manuscript in preparation).

There is also no evidence that indicates a "gradual displacement [of Neanderthals] to more marginal and less favorable environments, where their dwindling numbers would have suffered greater attrition from the vagaries of fluctuating climates and food supplies as well as disease"¹. Earlier, from the same evidence Stringer argued just the opposite², that Neanderthals persisted into the Upper Palaeolithic "... not just in backward or isolated areas either. Saint-Césaire is, in fact, situated in a region densely occupied during the Middle and Upper Palaeolithic." The marginalization interpretation would not seem to apply to the Neanderthals of western Europe, any more that it applies to the Neanderthals of western Asia, where Neanderthals and their so-called 'modern' contemporaries lived for a period estimated to be as long as 60,000 years, manufacturing identical industries and using the same technology to do so, applying their tools in similar ways (the microwear is identical), burying their dead with the same customs, hunting the same game species and butchering them the same way. Here, these two supposed human species are known to coexist with archaeologically indistinguishable adaptations for a very long time and there is no evidence of banging or whimpering. There simply are no data to support the assertion of a cultural, biological or ecological marginalization of Neanderthals.

Stringer and Grün suppose that Neanderthals may have been a different species, but then they assert that the species "were probably sufficiently closely related to allow hybridization", saying that "mitochondrial DNA studies have been used to suggest that there is no trace of a genetic input from Neanderthal females in recent European samples". (We know of no discovery of Neanderthal mitochondrial DNA.) To make sense of these contradictory claims one would have to assume that this new understanding of species is not based on reproductive isolation, and that the "hybridization" was between Neanderthal men and the women of the "anatomically modern" humans who (presumably) replaced them.

Stringer and Grün assert that "the simplistic equation of hominids and technologies in Europe has thus been abandoned". Yet they also state that the contemporaneity of 'modern' humans and Neanderthals is "established" by the earlier Aurignacian of western Europe, which is said to be "firmly associated with modern humans". But what new evidence supports such a "firm association"? Earlier, on the basis of the same data set, Stringer³ asserted just the opposite: "apart from the Saint-Césaire skeleton . . . we know nothing of the population of the beginning of the French Upper Palaeolithic, particularly those of the early Aurignacian". This contradiction remains unresolved. Moreover, in central Europe the earlier Aurignacian is associated with some fragmentary but diagnostic specimens that have several unique elements of Neanderthal morphology and no unique "modern" European features⁴. This is relevant to the question of who manufactured the earliest Aurignacian because the earliest 'modern' central European remains, such as Mladeč and Vogelherd, do not derive from the earliest Aurignacian levels. This does not prove that the earlier Aurignacian is associated with Neanderthals, but it surely suggests the need for more caution than expressed by Stringer and Grün. There is no evidence that indicates a "gradual displacement [of Neanderthals] to more marginal and less favourable environments, where their dwindling numbers would have suffered greater attrition from the vagaries of fluctuating climates and food supplies as well as disease"¹.

The replacement of Neanderthals by 'modern' populations in western Europe is treated as a fact. It is not. Rather than demonstrating the coexistence between Neanderthals and 'modern' humans, the new date reinforces the notion that in western Europe, as in central Europe, the actual Neanderthal remains always precede definitive remains of 'modern' populations. In the west these two now appear separated by a greater timespan. As far as central and western Europe are concerned, the new data provide more than sufficient reason to retract the declaration that "models centred on a direct ancestor-descendant relationship between Neanderthals and modern *Homo sapiens* must surely now be discarded"⁵.

MILFORD H. WOLPOFF

Department of Anthropology,
University of Michigan,
Ann Arbor, Michigan 48109, USA

DAVID W. FRAYER

Department of Anthropology,
University of Kansas,
Lawrence, Kansas 66045, USA

STRINGER REPLIES — The points raised by Wolpoff and Frayer make me wonder

whether we have made any progress since our last exchange with Wolpoff about the significance of Saint-Césaire, more than 10 years ago⁶.

The morphology of the Saint-Césaire fossil and of other actual or claimed late Neanderthals does not provide overall support for a model of local evolutionary continuity in Europe. Although it is true that in certain selected respects, such as nasal breadth and dental size, these specimens show a closer approach to later humans than do actual or presumed earlier Neanderthals, most features show stasis. In fact, for features related to midfacial projection (one of the best established Neanderthal apomorphies) Saint-Césaire deviates in the direction away from early modern European values.

Regarding Aurignacian hominids, Wolpoff and Frayer misquote the comment "firmly associated with modern humans", which referred to the industries which replaced the Châtelperronian and Szeletian, not the early Aurignacian, which is contemporaneous. It is surprising that the authors do not seem to be aware of a recent review listing possible and more definite hominids from the French Aurignacian (including Cro-Magnon), concluding that none shows significant Neanderthal features⁷. Regarding Vindija, I can only echo the words of Allsworth-Jones who stated "there are no convincing grounds for associating Neanderthal remains with the Aurignacian at Vindija"⁸. If, despite this, the presence of a bone point (a later Aurignacian artefact) is used to date these Neanderthal-like hominids, it would certainly seem to establish their broad contemporaneity with anatomically modern hominids such as those from Velika Pečina, Mladeč and Vogelherd⁹.

A further misunderstanding lies in the continuing confusion of different species concepts. As I have pointed out elsewhere^{10,11}, the morphological and phylogenetic species concepts I use do not depend on the presence or absence of hybridization and anyway, many closely related 'biological species' today are capable of hybridization. A demonstration that hybridization occurred between Neanderthals and early modern humans would not be proof that they were conspecific.

Wolpoff and Frayer answer our speculations about European events with far

more speculation about western Asia. If they are not willing to recognize a coexistence of Neanderthals and modern humans in Europe, it is difficult to see how they can be so sure of a coexistence (let alone one of 60,000 years) of the two types in western Asia from much less precise dating evidence.

Wolpoff and Frayer cite my 1988 News and Views piece on the thermoluminescence dating of the Qafzeh early modern *Homo sapiens* fossils in the Levant⁵. The estimated age of this material, since supported by other data¹², was more than 50,000 years greater than the date now obtained for Saint-Césaire. If this material does indeed indicate such an early appearance of the modern human morphology, the quote seems even more appropriate now than in 1988.

Beyond these arguments about interpreting the fossil and archaeological records, there is the matter of interpreting the dating evidence. There is complete statistical overlap between the thermoluminescence date for Saint-Césaire and radiocarbon dates for other Châtelperronian sites, and for numerous Aurignacian sites¹³. Another important issue was touched on by Mercier *et al.*¹⁴. This concerns the growing possibility, based on ice-core data and uranium series and thermoluminescence comparisons with ¹⁴C, that radiocarbon dates in the period around 30,000 radiocarbon years are significantly underestimating true ages. If this is confirmed by further work, it would not affect the established contemporaneity of the Châtelperronian and Aurignacian, but would provide even stronger evidence for a coexistence of late Neanderthals and early modern humans in Europe.

CHRIS STRINGER

Department of Palaeontology,
Natural History Museum,
London SW7 5BD, UK

1. Stringer, C. & Grün, R. *Nature* **351**, 701–702 (1991).
2. Stringer, C., Hublin, J.-J. & Vandermeersch, B. *The Origins of Modern Humans* (eds Smith, F. H. & Spencer, F.) 51–135 (Liss, New York, 1984).
3. Bouchud, J. *Bull. Soc. Et. Recherche Préhist. Les Eyzies* **15**, 1–9 (1966).
4. Wolpoff, M. H. *et al. Am. J. Phys. Anthropol.* **54**, 499–545 (1981).
5. Stringer, C. *Nature* **331**, 565–566 (1988).
6. Wolpoff, M. H., Stringer, C. B., Kruszynski, R. G. & Jacobi, R. M. *Nature* **289**, 823–824 (1981).
7. Gambier, D. *The Human Revolution* (eds Mellars, P. & Stringer, C. B.) 194–211 (Edinburgh University Press, Edinburgh, 1989).
8. Allsworth-Jones, P. *The Emergence of Modern Humans* (ed. Mellars, P.) 160–242 (Edinburgh University Press, Edinburgh, 1990).
9. Smith, F. H. *The Origins of Modern Humans* (eds Smith, F. H. & Spencer, F.) 137–209 (Liss, New York, 1984).
10. Stringer, C. B. & Andrews, P. *Science* **241**, 773–774 (1988).
11. Stringer, C. B. *Continuity or Replacement: Controversies in Homo sapiens Evolution* (eds Brauer, G. & Smith, F. H.) 9–24 (Balkema, Rotterdam, 1992).
12. Grün, R. & Stringer, C. B. *Archaeometry* **33**, 153–199 (1991).
13. Harrold, F. B. *The Human Revolution* (eds Mellars, P. & Stringer, C. B.) 677–713 (Edinburgh University Press, Edinburgh, 1989).
14. Mercier, N. *et al. Nature* **351**, 737–739 (1991).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

Napoleon's hat

SIR — Podsiadlowski *et al.*¹ propound an interesting double-cone model of the circumstellar nebula to explain the "Napoleon's hat" on the north side and other light echoes observed around supernova 1987A. Although they offer various predictions, they do not mention an important datum which was obtained some time ago: the spectroscopic radial-velocity gradient in the small central ring²⁻⁶. Their model cannot accommodate this datum.

The ring⁶, observed with the Hubble Space Telescope (HST) and from the ground, appears elliptical on the plane of the sky, with its major axis (1.7 arcsec) roughly east-west. Presumably it is a circular ring, inclined by about 43° so that either the north or the south point is closest to Earth. The radial-velocity gradient is small from east to west, but is positive from north to south. The spectroscopic data and images taken in best seeing (<1 arcsec)^{3,5} show that near the slit the surface brightness of the outer loops is small compared to that of the ring itself. Emission by the outer loops is indeed seen in the spectra but is weak. Therefore the observed velocity gradient is likely to be dominated by the velocity field of the ring itself. Supporting this is the fact that the observed gradient is coterminous with the ring's direct image⁵. The data are fitted well³⁻⁵ by a circular ring expanding radially from the progenitor star (or contracting onto it). With the reasonable assumption that the ring is expanding rather than contracting, the observers conclude that the north point is closest to Earth³⁻⁵.

Then the ring's plane is orthogonal to that shown for it in Fig. 2a of ref. 1. This is a serious discrepancy. In the geometrical and dynamical model of Podsiadlowski *et al.*, the ring was an equatorial belt or waist with symmetry axis roughly coinciding with that of the double cone. It was the remnant of an excretion disk produced in the orbital plane of the binary progenitor — a binary which also created the double cone. The double-cone geometry to explain Napoleon's hat, it seems, can be salvaged only by assuming either that the central ring somehow exists independently at the centre of the double-cone structure, with an axis orthogonal to the cone axis, or that the ring is now contracting onto the position of the binary progenitor which created it, rather than expanding. Neither assumption seems consistent with their dynamical model.

In a somewhat similar model by Wang and Mazzali⁷, the predicted surface brightness on the plane of the sky is symmetrical north-south (in their approximation), so the axis of the model can be flipped to match the ring's axis.

The same is true of a paper by Luo and McCray⁸, in which the ring's orientation is correct though mis-stated in the first paragraph. Neither paper^{7,8} contains an explanation of Napoleon's hat, and none of these models^{1,7,8} can do so when the ring is oriented correctly. As Kapteyn pointed out long ago, light echoes such as these contain much information about the circumstellar medium.

JAMES E. FELTEN
ELI DWEK

Code 685,
NASA Goddard Space Flight Center,
Greenbelt, Maryland 20771, USA

PODSIADLOWSKI *ET AL.* REPLY — Felten and Dwek raise an important point, but they oversimplify the situation and we do not agree with their conclusions concerning the Napoleon's hat nebula. Although the velocity gradient is observationally well established, its implication for the inferred orientation of the ring is model-dependent. The problem is that the spatial resolution of the spectroscopic observations is usually larger (in most cases much larger) than the north-south extent of the ring (~1"). Therefore, an overall velocity gradient would determine the orientation of the ring unambiguously only if there were no other significant contributions to the observed lines from elsewhere.

In contrast to Felten and Dwek, Wampler *et al.*⁹ find that the two outer loops surrounding the inner ring contribute significantly to the observed line fluxes (in O[III] the flux in the loops is one half the flux of the ring). Because of poor spatial resolution, the measured velocity gradient in a particular line can easily be dominated by the loops, whenever the flux in the loops is larger than about 10–20% and the characteristic radial velocity in the loops is larger than the radial velocity in the ring (the latter is model-dependent, but easily fulfilled in our model).

Therefore, an overall velocity gradient alone contains very little information about the orientation of the ring. The real information on the orientation can, however, be found in the detailed spectral line shapes (see, for example, ref. 5). Indeed, these may suggest an orientation of the ring different from the one we assumed. But a realistic model has to include the emission from the loops and the possible clumpiness of material in the ring as well as in the loops and should be able to reconcile some of the apparent discrepancies in the reported observations noted above.

A second even more important point is that, even if the ring's orientation were orthogonal to the orientation used in our figures, this would not at all affect our geometrical model for the Napoleon's hat nebula, which was the main

result of our paper (our geometrical model does not require any particular orientation of the ring). In particular, we would like to re-emphasize that our model has several unique features. In particular, we showed that the rim of Napoleon's hat has to have the same topology as the intersection of an elliptical cone with a paraboloid (this is entirely model-independent). Of course, Felten and Dwek are correct that a different orientation of the ring would make our suggested dynamical model for the whole nebula less attractive. Although it would not necessarily invalidate it, a different, probably more complicated, dynamical model would be required.

A real test of our model therefore has to come either from high-resolution images with the New Technology Telescope (NTT) or the HST or from detailed light echo studies (for example, the evolution of the inner light echo¹⁰ should directly map out the predicted cone structure). A recent picture¹¹ taken with the NTT shows that our emission model was not realistic. But because the particular emission model was chosen only for illustrative purposes, this is not important. What is important is the location of the northern end of the Napoleon's hat structure and the appearance of a southern cone (provided that it is not obscured). Another critical feature is the extension of the inner loops within the central ring. Close inspection of the HST image¹² shows and the earlier NTT image⁹ hints that the northern loop is closed, while the southern loop is not (both of which are consistent with our model). By now, our model predicts that the southern loop should also appear closed. A potential problem with this is that the HST may not have the dynamic range and the NTT the necessary resolution to resolve this question unambiguously.

PHILIPP PODSIADLOWSKI
ANDREW C. FABIAN
IAN R. STEVENS

*Institute of Astronomy,
University of Cambridge,
Madingley Rd,
Cambridge CB3 0HA, UK.*

- Podsiadlowski, P., Fabian, A. C. & Stevens, I. R. *Nature* **354**, 43–46 (1991).
- Wood, P. R. & Faulkner, D. J. *IAU Circ. No.* 4739 (1989).
- Crotts, A. P. S. in *Supernova 1987A and Other Supernovae* (eds Danziger, I. J. & Kjær, K.) 559–567 (ESO, Garching, 1991).
- Meikle, W. P. S. *et al.* in *Supernova 1987A and Other Supernovae* (eds Danziger, I. J. & Kjær, K.) 595–603 (ESO, Garching, 1991).
- Crotts, A. P. S. & Heathcote, S. R. *Nature* **350**, 683–685 (1991).
- Dwek, E. & Felten, J. E. *Astrophys. J.* **387**, 551 (1992).
- Wang, L. & Mazzali, P. A. *Nature* **355**, 58–61 (1992).
- Luo, D. & McCray, R. *Astrophys. J.* **379**, 659 (1991).
- Wampler, E. J. *et al.* *Astrophys. J.* **362**, L13 (1990).
- Crotts, A. P. S. & Kunkel, W. E. *Astrophys. J.* **366**, L73–L77 (1991).
- Gouffes, C. & Della Valle, M. *The Messenger, ESO* **66**, 34–35 (1991).
- Jacobson, P. *et al.* *Astrophys. J.* **369**, L63–L66 (1991).

Beyond good and evil

J. R. S. Fincham

The Genetic Revolution: Scientific Prospects and Public Perceptions. Edited by Bernard D. Davis. *Johns Hopkins University Press*: 1991. Pp. 294. £32.50, \$45 (hbk); £11.50, \$15.95 (pbk).

Biotechnics and Society: The Rise of Industrial Genetics. By Sheldon Krimsky. *Praeger*: 1991. Pp. 280. \$17.95 (pbk), \$47.95 (hbk).

Wonderwoman and Superman: The Ethics of Human Biotechnology. By John Harris. *Oxford University Press*: 1992. Pp. 271. £17.95, \$22.95.

DNA technology is remarkable not only for the great advances in biological knowledge that it has made possible but also for the disagreement and confusion that it has generated in the academic community and among articulate people in general. If any eventual consensus is to be reached we had better understand these differences and what lies behind them. Between them, these three books exhibit a fascinating variety of views ranging from hyperbole to scepticism and from enthusiasm to alarm.

Bernard Davis, who from the very first has been a militant anti-alarmist about genetic engineering, has assembled a panel of contributors to represent a variety of interests, excluding only the root-and-branch opposition. Five chapters review the outlook from various scientific viewpoints — microbes (A. M. Campbell), plants (C. J. Arntzen), animals (F. M. Loew), ecology (S. A. Levin), environmental protection (M. Mellon), molecular medicine (T. Friedmann), neuroscience (J. H. Schwartz) and evolutionary studies (F. M. Ayala). Davis reviews these chapters, disagreeing with his contributors as he thinks necessary. Three chapters deal respectively with public policy (A. Wildavsky), problems of regulation (H. I. Miller, an FDA regulator who thinks that no further regulation is necessary) and US regulatory law (R. B. Stewart, who outlines an extraordinarily tangled situation). H. Brooks and R. B. Johnson, both in the Harvard School of Government, conclude with a review of public policy issues, mainly those surrounding the possible dangers of environmental release, themes that recur throughout most of the book. Their chapter is balanced and optimistic, the authors envisaging an eventual consensus. "Fortunately", they say, "scientists are socialized with a strong ethic of building on objective evidence." It is good to see that ideal reasserted.

Objective evidence

Unfortunately, objective evidence about possible environmental hazards is not easily obtained. The basic trouble is that no hazard has yet been demonstrated. "Extrapolation from zero", to use Dav-

is's phrase, quoted by Wildavsky, is hardly possible. Ecologists seek to draw some lessons from experiences, sometimes unhappy, of introductions of alien plants and animals. Mellon thinks that the analogy between introduced species and genetically engineered organisms is rather a precise one. Levin refers to this as "probably the most contentious scientific issue" but then backs away with the words "Be that as it may . . .". I think he is right to be cautious in pushing the analogy. Organisms become invasive when they are no longer restrained by the environments in which they find themselves. In the case of an alien invasion, it is the organism's environment that changes, not its genetic equipment.

There seems to be agreement among the authors that it is the product of genetic change that matters, not the process that brought it about. But if this principle were really accepted, the result might be the extension of the same degree of concern to the products of conventional breeding, and that would seem contrary to experience and common sense. So genetic engineers will have to live with regulations based on the premise that there is something inherently risky in the process of *in vitro* DNA manipulation, even though most of them probably agree with Campbell that it is all a "chase after phantoms".

In the absence of agreed general principles, the commonly enunciated formula is "case-by-case evaluation", that is, judgement of each new product on its own merits. Levin, an ecologist, subscribes to this but also looks forward to the establishment — I would guess on the basis of common sense — of some general low-risk categories that will need only minimal scrutiny. Surely this is the realistic way forward, just as it was in dealing with earlier concerns about laboratory safety.

But Davis criticizes Levin for not abandoning the case-by-case approach altogether. Wildavsky, a political scientist, is even more outspoken. Commenting that in this area "few come out for the policy in which they believe nor believe the policy that they support", and despairing of attempts to extrapolate from zero, he advocates abandonment of

regulation in favour of just coping with problems as they arise, that is, of resilience rather than anticipation. Such a policy of rolling with the punch has an undoubted attraction, especially if one expects the punch, should it ever come, to be fairly light. But it is certainly a nonstarter in the climate of opinion documented in Sheldon Krimsky's book.

Commercial interests

Krimsky, a professor of social and environmental policy, seeks to interpret disputes in terms of social and cultural positions and commercial interests. In the earlier chapters, he deals with the rise of the biotechnology industry and the increasing commercial involvement of US academics. He clearly thinks that practising genetic engineers, and even university scientists in general, have largely forfeited their claim to act as impartial arbiters in risk assessment. Instead he looks to organized public-interest groups to curb the excesses of commercially driven technology. He does not actually applaud Jeremy Rifkin, accepting (on the authority of S. J. Gould) that he lacks scholarly integrity, but he admires him as an astute lawyer and an example to the public of how effective cleverly organized resistance can be.

Krimsky himself is no extremist opponent of "biotechnics". Indeed, on the penultimate page of his book he comes up with the statement: "The aggressive exploitation of genetic science for practical ends is by and large a healthy development." This endorsement, which comes as something of a relief after much that has gone before, suggests to me that he views genetic engineering less as a threat than as favourable ground on which to wage his larger campaign for "social guidance" of commercial technology of all kinds. As he readily admits, he is not disinterested. The bias is apparent, but a great deal of valuable historical record, in the form of bibliography, factual commentary and apt quotation of contrasting opinions, does show through. Social guidance is indeed a big issue. One cannot disagree with the principle, but its benefits will depend on the acquisition by the public of a degree of immunity to sensationalist journalism.

Both Friedmann (in Davis's book) and Krimsky discuss human gene therapy, foreseeing real benefits but making some reservations. Friedmann remarks that "fortunately there are limitations to the extent that we can redesign ourselves genetically". That seems to me to be a nice understatement, but John Harris, writing as a professor of applied philosophy, has no such timidity. Having swallowed whole, as it seems, all the journalistic hype that has bedevilled public understanding of the field, he has little

feeling for biological complexity or logistical difficulties. Take, for example, this statement: "When the human gene project is completed, or very soon thereafter, the process of single cell biopsy will enable the complete 'book of life' for each individual to be compiled and of course to be read." The mind boggles.

Harris takes some very clearly argued if provocative moral positions. Briefly summarized, they are that *prepersons* — that is that stage of human life that in his view starts (rather strangely) at germ-cell formation and ends at some significant (but unspecified) time after birth — have no rights in themselves, but that after they have become *persons* they all have the right to the best life attainable. He believes that genetic therapy or enhancement (including germ-line manipulation) should be provided for as many as possi-

ble, even if for practical reasons (a flash of realism), it cannot be made available to everyone who might benefit from it. He is strong on consumer choice. What could be wrong about parents being able to choose curly hair or brown eyes for their little daughter? Why should not people be "their own fairy godmothers"? To which one can only reply that, thought experiments aside, anyone claiming to have a magic wand is certainly a quack from whom potential people should be protected.

This highly individual book is of some intellectual interest. But one can only hope that the naive public does not take it all as referring to the real world. □

J. R. S. Fincham is in the Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

In defence of alchemy

Maurice Crosland

The French Paracelsians: The Chemical Challenge to Medical and Scientific Tradition in Early Modern France. By Allen G. Debus. Cambridge University Press: 1992. Pp. 247. £40, \$59.95.

ALLEN Debus has devoted most of his professional life to the study of the "Swiss-German firebrand" Paracelsus (1493–1541) and his followers. In the 1970s, Debus began to assume some prominence among a group of scholars reacting against a traditional history of science that presented a triumphant rational progress in the physical sciences from Galileo to Newton and beyond. A study of Newton's manuscript notes had revealed that Newton had an extensive interest in alchemical matters, so there seemed to be some justification for adjusting historiographical perspectives. Were there not other paths that early modern science might have followed? Should alchemy simply be dismissed as irrational? Surely some attention should be given to the less well-known undercurrents in the early history of science. Nor should they necessarily be regarded in a negative way. Paracelsus had presented alchemy rather than mechanics as the most fundamental science, and paracelsian cosmology and natural philosophy represented a powerful alternative tradition within the emerging science of the late sixteenth and early seventeenth centuries. Debus has felt that he had to compensate for the later comparative neglect of this alternative tradition by making great claims for its importance.

Debus begins by summarizing some of his earlier work for the benefit of new

readers. He then breaks new ground by going beyond the seventeenth century and by focusing on France, which was after all, with Britain, among the most important centres for the emergence of modern science. Much of the book is concerned with the conflict between the followers of Paracelsus, who advocated new chemical remedies in medicine, and



Paracelsus — brilliant and abrasive alchemist.

the followers of Galen, here represented by the ultra-conservative Paris Faculty of Medicine.

The approach is a bibliographical one. We pass in chronological order from one author to the next and, because there is sometimes a danger of simply listing publications and describing the positions of their authors, Debus very sensibly ends each chapter with a section entitled "Conclusion". We must give him credit for having unearthed many books that, in his own words, "have lain practically untouched in libraries for hundreds of years", yet one may ask whether he has done more than simply wipe away the dust of centuries.

Debus argues convincingly that it was through its utility in medicine that che-

mistry gradually gained institutional recognition. Debates on chemical medicines continued fiercely in the seventeenth century, but after this his conclusions are more controversial. Just because a handful of books in the eighteenth century continued to discuss alchemy, Debus concludes that "the eighteenth century was a period of intense interest in alchemy" and that even then the search for the philosopher's stone was actively pursued. Many would question such claims.

In "the age of reason", alchemy seemed to most intellectuals and men of science an aspect of history best forgotten. Of course, the psychological appeal of making gold is timeless, but the concerns of the age were elsewhere. This could easily be shown by a quantitative analysis of the subject matter of the whole range of eighteenth-century publications. We also need to know more not only about authors but about audiences. The few pages devoted to Geoffroy in the *Mémoires* of the Paris Academy of Sciences (1722) to what was actually a refutation of alchemy, was one of the last occasions that the subject was mentioned in that company.

Debus confirms that in the eighteenth century the requirement for a medical justification for chemistry was diminishing in importance. Of necessity, the story of the paracelsians must eventually be one of decline. One or two readers might have hoped for some grand denouement, for example some evidence that Lavoisier, the founder of the new chemistry, was all the time a secret paracelsian, or at least — like Newton — had some association with alchemy. Any evidence of such influence will have to be left to the research of a future generation of historians.

Debus has too great a respect for printed texts to invent such a connection. Yet if one were to go beyond these texts and their authors, one might be in a better position to appreciate how they fitted into the whole historical picture. It is remarkable that the meticulous research on which this book is based does not seem to have included even one visit to a library in France. Nor is there any real engagement with the ideas of the Enlightenment. This weakens the claim of the book to be a contribution to intellectual history. Yet some of the ideas that it discusses will have to be incorporated in any future synthesis. Debus has made an important contribution to the history of pharmacy and to some aspects of the early history of chemistry and of medicine. □

Maurice Crosland is in the Unit for the History, Philosophy and Social Relations of Science, Physics Building, University of Kent, Canterbury CT2 7NR, UK.

A case of battle fatigue

H. M. Collins

War in the Age of Intelligent Machines.

By Manuel De Landa. *Zone Books*: 1992. Pp. 271. £14.95, \$22.95 (pbk); £29.75, \$44.50 (hbk). Distributed by MIT Press.

THE first half of this book is an intriguing history of control over warfare. De Landa questions whether the point of control over battles should lie in the centre, demanding a rigidly organized, unthinking, 'clockwork' soldiery. Or is war more successful when the centre determines the overall goal, with small but powerful platoon units deciding how best to attain their own objectives? De Landa shows that the appropriate answer depends on the economics, politics and technology of war.

What can be said in favour of centrally organized, closely coordinated bodies of troops? For one, desertion can be curbed with rigid plans of battle that keep troops close to one another and to the enemy. And inaccurate musketry requires volley fire if it is to be effective, again demanding central organization. On the other hand, the invention of the national army allows the replacement of scarce, expensive and potentially disloyal mercenaries with an endless reservoir of cheap patriots, thereby reducing the need for central control. And with the rifled barrel and the cone-shaped bullet comes improved accuracy, which lessens the importance of volley fire. Radio enables geographically dispersed and disparate units to coordinate their actions. With infantry working with artillery and aircraft, the blitzkrieg, as opposed to the organized set-piece confrontation, is the perfect form of war in the age of radio.

All this is fascinating, particularly as war can represent any organized human endeavour. But of course, nothing in real history is ever quite so straightforward. Was not the decisive Allied victory over the Germans at El Alamein in 1942 a set-piece victory in the age of blitzkrieg? The hard part of writing such a history is making sense of the interactions of humans with new technological developments. De Landa's view gives a fixed meaning to the barrel and the bullet, with generals either grasping the opportunities or failing to see them. This is all too simple; it is a view that comes with perfect hindsight.

De Landa's description of war planning after the Second World War pits Rand think-tank mathematicians against real soldiers. He emphasizes the 'friction' of battle, saying that "no plan

survives contact with enemy". War games, he argues, deal with a dehumanized battlefield, the mathematical descriptions of weapons being determined not by experience in the hands of exhausted and terrified fighters but by exaggerated claims originating from interservice rivalry. Even in war games, it turns out that humans will not press the imaginary nuclear button. De Landa therefore argues that nuclear planning "takes the humans out of the loop", allowing computers to be pitted against one another in an effort to form a "rational" plan for the destruction of the planet. The removal of humans from the loop is the topic that drives the second half of the book. The process started with automated gunnery in the Second World War, has now moved to antiradar missiles that select and home onto their targets, and threatens to dispense with people entirely as artificial intelligence takes over.

Here De Landa's book is weak. Indeed, one gets the impression that the manuscript was dusted off in the aftermath of the Gulf War and given a sexy title that does not really fit its contents. Despite the promise on the back cover, the Gulf War is not mentioned and the discussion of artificial intelligence is dated. Neural networks receive no coverage, with expert systems, parallel processing and hypertext described as if they were the latest thing. Writing about artificial intelligence is a hazardous business. Everyone believes a version of the myth that "Hubert Dreyfus predicted that computers would never play chess and he was soon proved wrong" (this myth is not quite true, by the way) and authors are reluctant to be too critical of the 'latest developments', even when these developments have already had their day in terms of hype. Admittedly, De Landa seems to understand the problem of artificial intelligence. He knows that an autonomous battlefield robot with a licence to kill will have to make inductive generalizations, and he knows that if it is to induce as we do, it will have to know everything we know. But he does not seem sure whether this state of affairs is "just around the corner".

The second half of the book, then, loses its way. At the same time, the writing loses its style, becoming full of ambivalence, repetition, split infinitives and irrelevant overworked metaphors. Unfortunately, the subject of the second half is what the book is supposed to be all about. One could learn a lot about the history of warfare from the book, but nothing at all about war in the age of intelligent machines. □

H. M. Collins is in the School of Social Sciences, University of Bath, Claverton Down, Bath BA2 7AY, UK.

Astro-design

R. Hanbury Brown

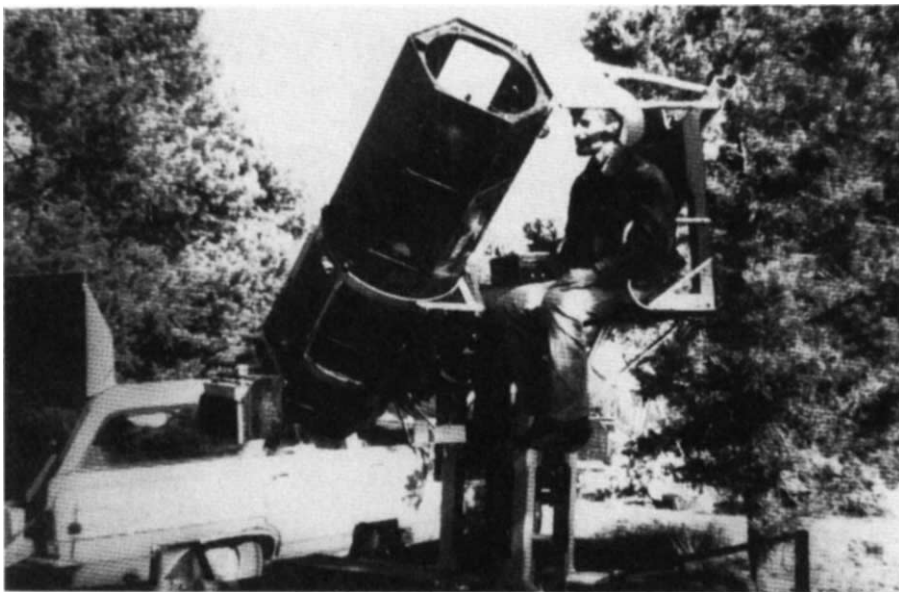
Unusual Telescopes. By Peter L. Manly. *Cambridge University Press*: 1992. Pp. 221. £19.95, \$39.95.

ONE of the things I like doing in museums, particularly motor museums, is looking at the variety of ways in which different designers have solved the same problem. In the struggle for survival of the fittest, only a few of the solutions ever reached conventional design: many were simply not the best way of doing the job, while some arrived before their time, only later being resurrected. It is the same with telescopes, and to read Peter Manly's book is the next best thing to going round a museum. The book is clearly written, well illustrated and both instructive and entertaining.

Most of the book is devoted to the principal problems that must be solved in building a telescope, including those of the optical system, mount, materials and drive. For each, we are told about some of the many ways in which the problems have been tackled, the emphasis being on unconventional solutions. Most, but not all, of the unconventional designs have been made by amateur astronomers; broadly speaking, these people build smaller equipment than the professionals, usually spend their own money and can take greater risks.

The longest chapter is on optics. Manly covers the problems of making mirrors out of all sorts of material, not just glass, which is not sufficiently unusual, but also rock, clay, metal, plastic, liquids, membranes, gases and so on. I particularly liked his well-illustrated discussion of optical systems with different numbers of optical elements. He tells us, for example, about many of the ingenious ways in which people have tackled the problem of collecting the light from a telescope without obstructing the primary mirror, examples of which are the schiefspiegler and off-axis Cassegrain telescopes.

There are also problems associated with mounting a telescope. Although most telescopes have two axes, there have been telescopes made from zero to four moving axes, as well as those that move freely in any direction. Manly gives examples of all of these, some of them being "blatantly odd", but others, such as the single-axis mount of the transit telescope at the US Naval Observatory, being conventional. The chapter called "Strange Drives" lives up to its name. The problems of driving a telescope smoothly and at the correct rate have been solved mainly by the use of ingenious gear technology, and Manly provides examples of makeshift drives



Riding high — sometimes it is easier to attach one's self to the end of one's telescope than to follow the eyepiece round. Pictured here is Pierre Schwaar's merry-go-round telescope.

made from such bits and pieces as a Meccano set, a fishing reel and even a water clock. Heath Robinson would have enjoyed this chapter.

Another intriguing problem is whether the eyepiece should move or be stationary. When looking up at the immense columns of masonry that support the vast tube of the 1.83-metre (72-inch) telescope built by William Parsons, the Third Earl of Rosse, at Birr Castle, Parsonstown, in the mid-nineteenth century, I have sometimes tried to imagine what it was like on a windy night to follow in the dark the moving eyepiece at its newtonian focus. I was pleased to find that Manly uses this great telescope, the Leviathan, to illustrate the difficulties of a moving eyepiece: by any standards it was blatantly odd. He might well have also mentioned the observer's cage at the prime focus of the 200-inch telescope at Mount Palomar — by modern standards that is certainly odd. How to provide a moving telescope with a stationary eyepiece is a classical problem that has been tackled in many ways. Manly tells us about some of the solutions, among them the Springfield mount, which is basically a newtonian telescope mounted at the eyepiece end.

Although this book is a collection of oddities, it is not just a rag-bag. Manly tries to fit his examples into a coherent framework structured around the problems of designing telescopes and makes copious references to articles giving more information. To my mind, he would have done better to restrict his coverage to optical telescopes instead of including instruments designed for radio, infrared and cosmic rays. For example, he does not do justice to the oddities and problems in designing radiotelescopes — that would need a book in itself.

Many astronomers, in fact many scien-

tists, are insufficiently aware that it is the progress of instrument design that spearheads the progress of their science. Too many of them do not know how their instruments are made or what is inside the black boxes that they use; it would do them good to read a book like this. □

R. Hanbury Brown, emeritus professor of physics at the University of Sydney, is at the White Cottage, Penton Mewsey, Andover, Hampshire SP11 0RQ, UK.

Science and the underdog

John W. Galloway

Cancer Selection: The New Theory of Evolution. By James Graham. *Aculeus*, PO Box 142, Lexington, Virginia 24450, USA: 1992. Pp. 213. \$20.

WE have a good idea of the sorts of molecular events that underlie the initiation and growth of cancers. They all involve damage to a narrow spectrum of a few dozen genes (or the neutralization of the proteins for which they code). And all these genes concern the regulation of DNA replication and cell division, or differentiation.

Do we have a comparable picture of the molecular basis of evolution? James Graham thinks we do. He believes it is this self-same set of genes. Mr Graham, an avowed amateur scientist, has a bone to pick with evolutionists. He does not think that the theory of evolution in itself explains the appearance of complex organisms such as people, although it is OK for simple things such as plants and colonial animals. I think I agree with him. Evolution 'by natural selection'

does not explicitly seem to explain complexity (notwithstanding the eloquence of people such as Richard Dawkins). Plenty of simple things are extremely successful. Graham suggests that complexity can be roughly measured by the size of an organism and the number of different cell types it contains. And, because many cancers probably result from the failure of proliferation and differentiation mechanisms, he is able to argue, as he already has in the *Journal of Theoretical Biology*, that cancer might drive such mechanisms to being more varied — and organisms to becoming more complex.

That cancer is the result of perturbations in the molecular mechanisms underlying development presumably goes without saying. The idea that any cell embarking on a cancerous path is part of an evolutionary experiment seems self-evident; the suggestion that much diversity (within the immune system for example) is the result of the evolutionary pressure of infections and other diseases in the past can hardly be argued with.

Finally, it seems fairly well established that, for animal cells, differentiation and cell division act against one another, with cancer as a 'tug of war' between the two processes. Graham would thus have us believe that diversification of differentiation mechanisms is the result of a defensive response to uncontrolled cell division. I, at least, like the idea.

Unfortunately, Graham spoils his good idea by his advocacy of it. He is simply too self-indulgent, nowhere more obviously than in his attacks on those who have not accepted his theory and those who he thinks would not have accepted it, had they had a chance. These include a variety of well-known evolutionists of all possible stripes — Darwin, Stephen Jay Gould, Lysenko, Marx. (Indeed, he seems to have detected a marxist conspiracy on the loose, trying to suppress his ideas.) Calling them all crackpots and questioning their integrity can be a bit wearisome unless it is done with imagination.

Graham's abuse of scientists tells us little about them but much about him; it is the language of the underdog. Indeed, my advice to him would be to stop writing for a bit and start reading (how about Gould's *Wonderful Life*?) — and, of course, to become more philosophical. I am a great believer in everyone, not just the professionals, joining the crew of the good ship Science. But that does not give Graham an automatic right for his ideas to be enthusiastically piped aboard. □

John W. Galloway is at the Nuffield Foundation, 28 Bedford Square, London WC1B 3EG, UK.

Six years of the fifth force

Ephraim Fischbach & Carrick Talmadge

The enunciation of the 'fifth force' hypothesis in 1986 spawned a generation of experiments searching for deviations from newtonian gravity. Although no compelling evidence for any new weak forces has emerged in the past six years, the searches for anomalous gravitational effects have produced a large number of important experimental and theoretical results.

THE suggestion roughly six years ago of a possible gravity-like 'fifth' force¹ of nature called attention to the fact that gravity, itself the 'oldest' of the known forces, was in some ways also the least well understood. As we discuss below, a fifth force coexisting with conventional gravity would lead to a net interaction between macroscopic objects which showed small deviations from the behaviour expected from the classical newtonian inverse-square law of gravity (see equations (2) and (3) below). Evidence for such non-newtonian behaviour in an appropriate system could thus be a sign of a new fundamental force, and this idea has provided part of the impetus for the current efforts^{2,3} to re-examine the experimental support for newtonian gravity over various distance scales. Although no compelling experimental evidence for any anomalous results has yet surfaced, there has nonetheless been a dividend from the intense effort of the past few years. This is the substantial improvement in the precision with which newtonian gravity can be claimed to be experimentally supported. Moreover, the search for non-newtonian effects continues, spurred on both by remarkable advances in experimental techniques and by an improved understanding of the theoretical basis for such effects. Specifically, we now recognize that many models predict the existence of new intermediate-range forces, and relate the properties of these forces to the behaviour of physics at the Planck (energy) scale.

It is important to bear in mind that although we will focus here primarily on what has been learned in this field since the fifth force was proposed in 1986, this proposal was itself a natural outgrowth of many earlier experimental and theoretical studies of non-newtonian gravity². In the 1970s interest was stimulated by the work of Fujii⁴ and others^{5,6} who argued that various theoretical models led naturally to a new intermediate-range force, whose strength was comparable to (or perhaps slightly weaker than) gravity. Although Fujii's theory was based on detailed dynamical considerations, at its core are a few fundamental observations which have helped to motivate much of the subsequent work. We begin by noting that there are two natural mass scales in physics, defined by $m_N \equiv 1 \text{ GeV}/c^2$ and $M_{\text{Planck}} = (\hbar c/G)^{1/2} \approx 10^{19} \text{ GeV}/c^2$, where m_N is the nucleon mass, $\hbar$ is Planck's constant, c is the speed of light, and $G = (6.67259 \pm 0.00085) \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ is the newtonian gravitational constant. The ratio of these scales, $f \equiv m_N/M_{\text{Planck}} \approx 10^{-19}$, introduces a new small dimensionless constant into physics. This constant may have some dynamical significance if it determines the coupling strength of a new field to matter, just as the electric charge e ($e^2/\hbar c \equiv \alpha_{\text{em}} \approx 1/137$) determines the strength of the coupling of photons to matter. If this were the case, then the analogue of α_{em} for this field would be $f^2/\hbar c \approx 10^{-38}$, which corresponds to a force of gravitational strength ($f^2/Gm_N^2 = 1$). The second observation is that in various theories the product $\sqrt{(f^2/\hbar c)} m_N \equiv \mu = 10^{-10} \text{ eV}/c^2$ determines the mass of the new field. As the Compton wavelength (which characterizes the distance over which a force acts) associated with this field would be $\lambda = \hbar/\mu c \approx 2,000 \text{ m}$, this combination of values for the parameters f and μ could describe a new intermediate-range field of gravitational strength.

To understand how such a field could be detected experimentally, we note that the potential energy $V_5(r)$ describing the

interaction of two point masses m_i and m_j through this new force would be given by

$$V_5(r) = -\alpha \frac{Gm_i m_j}{r} e^{-r/\lambda} \quad (1)$$

where by convention the sign has been chosen to make $V_5(r)$ correspond to an attractive force for $\alpha > 0$. Here $r = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between the masses, and α is a dimensionless constant proportional to f^2 which determines the strength of the new force (relative to gravity). Because α can be either positive or negative in various theoretical models (corresponding to a force which is attractive or repulsive), the sign of α carries important physical information. For example, a repulsive force typically arises from the exchange of a vector (spin-1) meson between nucleons or electrons. Theories of such forces lead us to expect that the magnitudes of the effects they produce necessarily depend on the chemical compositions of the test masses. By contrast, an attractive force between like objects arises from the exchange of scalar (spin-0) and/or tensor (spin-2) fields, and these can lead to deviations from the $1/r^2$ law that are independent of the compositions of the test masses. It follows that an experimental determination of the sign of α can be used to discriminate among different theories of the fifth force. Returning to equation (1), we note that the sum of $V_5(r)$ and the usual newtonian potential $V_N(r)$ can be written as

$$\begin{aligned} V(r) = V_N(r) + V_5(r) &= -\frac{Gm_i m_j}{r} - \alpha \frac{Gm_i m_j}{r} e^{-r/\lambda} \\ &= -\frac{Gm_i m_j}{r} (1 + \alpha e^{-r/\lambda}) \end{aligned} \quad (2)$$

In the form of equation (2), the presence of the new interaction described by $V_5(r)$ manifests itself as an apparent deviation from the usual $1/r$ newtonian potential. In practice nearly all experiments measure a force (or acceleration) rather than a potential, and from equation (2) we find for this force

$$\begin{aligned} \mathbf{F}(r) &= -\nabla V(r) = -G(r) \frac{m_i m_j \hat{\mathbf{r}}}{r^2} \\ G(r) &= G_\infty [1 + \alpha(1 + r/\lambda) e^{-r/\lambda}] \end{aligned} \quad (3)$$

In equation (3) we have written the newtonian constant as G_∞ to indicate that this is the constant that determines the strength of the interaction in the limit $r \rightarrow \infty$. This serves to distinguish G_∞ from the constant $G_0 = G_\infty(1 + \alpha)$, which characterizes the interaction strength when $r/\lambda \ll 1$. (The latter regime is often referred to as the 'laboratory scale' on the assumption that λ is larger than the typical separation of test masses in laboratory experiments.) It follows from equation (3) that any deviation of $G(r)$ from G_∞ leads to a breakdown of the usual $1/r^2$ force law, and can be interpreted as evidence for a new force.

Elementary phenomenology

The work of Fujii and others, which led to the phenomenological framework given in equations (1)–(3), motivated further experimental and theoretical work, as discussed in ref. 7. We see from

equation (3) that every experimental result implies some constraint on the parameters α and λ which define the putative non-newtonian interaction. For technical reasons, the limits on $\alpha = \alpha(\lambda)$ implied by a given experiment are most stringent for values of λ that are comparable to the characteristic separation r of the test masses in that experiment. It follows that no single experiment can provide useful limits on $\alpha(\lambda)$ for all values of λ . Rather, a collection of experiments is needed, each optimized to measure $\alpha(\lambda)$ near a selected value of λ . By 1981 a number of such results were available, and these were collected together in an important paper by Gibbons and Whiting⁸ (hereafter referred to as GW).

The constraints on α and λ implied by the data available to GW are shown by the dark shaded area in Fig. 1. In this and similar figures (see, for example, refs 9, 10), the shaded region in the α - λ plane is excluded by the data at the two-standard-deviation (2σ) level. The GW paper drew attention to the region ($10\text{ m} \leq \lambda \leq 10^3\text{ m}$) where the limits on α (and hence on the validity of newtonian gravity) were very poor, as can be seen from Fig. 1. This gap in our knowledge simply reflects the difficulty in designing experiments in which the separation of the test masses is of order 10 – 10^3 m . Typically information on gravity for such mass separations comes from geophysical experiments, and hence this range of distances is often referred to as the 'geophysical window'.

Stacey and collaborators^{11–14} undertook to explore the 'geophysical window' by reviving the Airy method of determining the newtonian gravitational constant G . A schematic outline of the Airy method is shown in Fig. 2 for an idealized model of a spherical non-rotating Earth.¹⁵ Suppose we compare the acceleration $g(0)$ of a test mass at the surface of the Earth with its acceleration $g(z)$ at a depth z below the surface. (The test mass in each case is simply the proof mass in a standard gravimeter, usually of the LaCoste-Romberg type.) It is straightforward to see that there are two different effects that contribute to $\Delta g(z) \equiv g(z) - g(0)$, and that these have opposite signs. The first is the free-air gradient, which represents the increase in $g(z)$ arising from the circumstance that the test mass at z is only a distance $(R - z)$ from the centre of the inner mass, rather than at R as would be the case at the surface (see Fig. 2). The second effect, known as the double-Bouguer term, accounts for the fact that at a depth z , $g(z)$ will decrease because the test object is now attracted to the centre of the Earth by a smaller total mass $(M - \Delta M)$. (Recall that if Newton's $1/r^2$ law of gravity holds, then the shell of matter of thickness z does not contribute to $g(z)$.) From Fig. 2 and the preceding discussion

we see that in this simplified model $\Delta g(z)$ is given by

$$\Delta g(z) = g(z) - g(0) = \frac{G(M - \Delta M)}{(R - z)^2} - \frac{GM}{R^2} \quad (4)$$

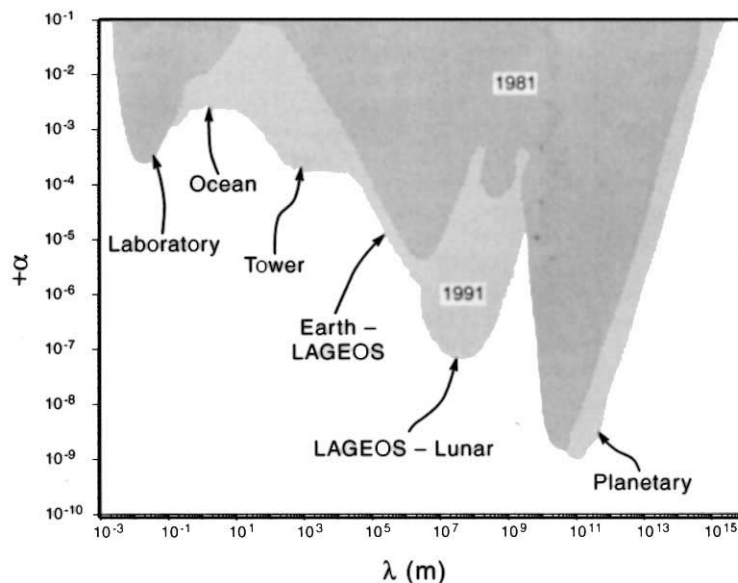
$$\approx \frac{2g(0)z}{R} - 4\pi G\bar{\rho}z$$

where $\bar{\rho}$ is the density of the material in the shell, and $g(0) = GM/R^2$. It follows from equation (4) that G can be determined over geophysical scales by a series of local measurements to determine $g(z)$, $g(0)$ and $\bar{\rho}$, and this constitutes the Airy method. In practice equation (4) must be modified to include the effects of the Earth's rotation and ellipticity^{11–15}. The original analysis by Stacey and coworkers¹³ in the early 1980s in fact found an anomalously high value of G : their best value, obtained from a mine at Mount Hilton in Australia, was $G = (6.720 \pm 0.024) \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ which was larger than the laboratory value $G = (6.67259 \pm 0.00085) \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ by roughly twice their maximum admitted error.

We note for later purposes that as $G(r)$ always appears in combination with a mass such as m_i , $G(r)$ cannot be determined separately unless m_i is independently known. In the Airy method the appropriate combination is $G(r)\Delta M$, and as ΔM can be determined from measurements of the local density $\bar{\rho}$, $G(r)$ can be inferred for values of r comparable to the depth of the measurement. But in some situations, such as tower experiments and tests involving planetary motion (see below), the appropriate masses cannot be determined independently. In these other cases G cannot be obtained in absolute terms, and only the variation of $G(r)$ with r can be measured.

During the same period of the early 1980s, Aronson *et al.*¹⁶ analysed high-energy Fermilab data on the K^0 - $\bar{K}^0$ system to look for anomalous effects that might be associated with the existence of a new force. They used neutral kaons because the small mass difference $\Delta m = m_L - m_S = 4 \times 10^{-6} \text{ eV}/c^2$ of the weak interaction eigenstates K_L and K_S makes this system extremely sensitive to small external influences^{17,18}. The presence of a new weak force would show up as an apparent dependence of Δm on the velocity v of the kaons (or their energy) in the laboratory, because the strength of an external field as seen in the rest frame of the kaons depends on v through the Lorentz transformations. Aronson *et al.*¹⁶ also noted that similar velocity-dependent effects could show up in the other fundamental parameters of the K^0 - $\bar{K}^0$ system, such as the lifetimes τ_L and τ_S of K_L and K_S , and CP -violating parameter $\eta_{\pm} \equiv \text{amplitude}(K_L \rightarrow \pi^+ \pi^-)/\text{amplitude}(K_S \rightarrow \pi^+ \pi^-)$. Motivated by these theoretical

FIG. 1 Constraints on the coupling constant α as a function of the range λ from composition-independent experiments. The dark shaded area indicates the status as of 1981, and the lighter region the current limits. For references to the earlier experiments which contribute to the curves, see refs 9, 10.



considerations, they reanalysed data from a series of Fermilab experiments which determined the K^0 - $\bar{K}^0$ parameters for kaons in the energy range 30–130 GeV. They found evidence at a marginally interesting ($\approx 3\sigma$) level for an energy variation of the kaon parameters, particularly for the phase $\phi_\pm$ of $\eta_\pm$, and this provided at that time another hint for a possible new force. We will return later to discuss the present status of kaon experiments.

The 'fifth force' hypothesis^{1,7} in 1986 arose from the recognition that the anomalies reported by both Stacey *et al.*^{11–14} and Aronson *et al.*¹⁶ could be explained in terms of the same new force, but only if it had properties that were somewhat different from those of $V_5(r)$ in equation (2). Additionally, the requisite force would be capable of producing anomalies in the classic experiment of Eötvös, Pekár and Fekete (EPF)^{19,20}, which compared the accelerations of different pairs of materials to the Earth. This motivated a reanalysis^{1,7} of the Eötvös experiment, from which two surprising observations emerged. First, the results of EPF indicated that for some pairs of materials, the acceleration differences for the two samples were relatively large (5σ in the case of copper against H_2O). Second, these acceleration differences fitted a pattern that was compatible with the existence of a new 'fifth force'. This observation, coupled with the suggestion that the same force might also account for the earlier anomalies^{13,16} helped to revive interest in non-newtonian gravity.

The proposed fifth force departed from the earlier work of Fujii by introducing a repulsive interaction proportional to the hypercharge of a test object $Y = B + S$ (B is baryon number, S is strangeness), rather than to its mass as in equation (2). For ordinary bulk matter $B = N + Z$ and $S = 0$, where N and Z respectively denote the number of neutrons and protons in a sample, and hence Y and B can be used interchangeably. But for certain elementary particles such as kaons (specifically K^0 and its antiparticle $\bar{K}^0$), $B = 0$ and $S \neq 0$. Hence by selecting Y as the charge we allow interactions not only between samples of ordinary bulk matter, but also between bulk matter and kaons. Such an interaction could account for the anomalies reported by Aronson *et al.*¹⁶, and for this reason hypercharge emerged as the natural choice for the charge in the original fifth force

theory. As our primary focus here is on macroscopic experiments, we can set $Y = B$ and write $V_5(r)$ in the form

$$V_5(r) = +f^2 \frac{B_i B_j}{r} e^{-r/\lambda} \quad (5)$$

where f is a new fundamental constant. It is straightforward to show that when $V_5(r)$ in equation (5) is combined with the newtonian potential $V_N(r)$, the resulting expression for $V(r)$ has the form given in equation (2), but with α replaced by α_{ij}

$$\alpha_{ij} = -\xi \left(\frac{B_i}{\mu_i} \right) \left(\frac{B_j}{\mu_j} \right) \quad (6)$$

where $\xi = f^2 / G_\infty m_H^2$, $\mu_i = m_i / m_H$ and $m_H = m({}_1H^1)$. As (B_i / μ_i) and (B_j / μ_j) vary from one material to another, α_{ij} is a function of the compositions of the interacting materials. It then follows from equations (3) and (6) that the accelerations of two samples j and j' towards a common source i would depend on the constants α_{ij} and $\alpha_{ij'}$ respectively, and these are generally unequal. If i denotes the Earth then the accelerations of j and j' towards the Earth would depend on their compositions, and this could be tested by simultaneously dropping two dissimilar objects as Galileo is supposed to have done^{21–23}. Such an experiment, although relatively simple in principle, is extremely difficult in practice and has been carried out to high precision only recently^{24,25}, as we discuss below. Eötvös recognized, however, that an equivalent experiment could be done by suspending the two samples j and j' from a common bar, and searching for a torque on the bar arising from the differential force exerted on the samples by the Earth. Subsequent experiments^{26–28} used the Sun as a source, and these set limits on possible new weak forces for which λ is of order 1 AU.

It follows that in the model considered here, the acceleration difference $\Delta a_{jj'}$ of j and j' towards the Earth is given by

$$\Delta a_{jj'} = \xi \left(\frac{B}{\mu} \right)_\oplus \left[\left(\frac{B}{\mu} \right)_j - \left(\frac{B}{\mu} \right)_{j'} \right] \mathcal{F} \quad (7)$$

where $\mathcal{F}$ is the field strength (in units of acceleration) of the source, which in this case would be the Earth (denoted by $\oplus$). As in equation (12) below, B can be replaced by a more general charge Q , so that $(B/\mu)_\oplus \rightarrow (Q/\mu)_\oplus$ and so on. It then follows that the experimentally determined sign of $\Delta a_{jj'}$ (that is, whether j falls faster than j' or vice versa) depends on the signs of the factors $(Q/\mu)_\oplus$ and $[(Q/\mu)_j - (Q/\mu)_{j'}]$, and on the direction of $\mathcal{F}$. It should be emphasized that for a force of finite range the effective values of $(Q/\mu)_\oplus$ and $\mathcal{F}$ are actually determined by the local chemical composition and the local matter distribution, respectively, and can vary from one location to another. As this information for the site of the original EPF experiment is not accurately known, the sign and magnitude of α implied by this experiment cannot now be established unambiguously.

From equation (3) we see that with $\alpha \rightarrow \alpha_{ij}$, an intermediate-range weak force can be detected through a modification of the inverse-square law ($G(r) \neq \text{constant}$) and/or through a composition-dependence of the net acceleration ($\alpha_{ij} \neq \text{constant}$). (The latter effect is often referred to as a violation of the weak equivalence principle or of the universality of free-fall.) It is evident from equation (3) that both effects are generally present in a given experiment. In practice, however, experiments are usually designed to isolate one or the other of these signals for a new force, and we will therefore focus on each of these classes of experiments separately.

Composition-independent experiments

We first consider the recent progress made in tests of the inverse-square law. Returning to Fig. 1, we see that the limits on the strength $\xi \equiv -\alpha$ of a possible non-newtonian component have improved substantially over several distance scales. Most notably the geophysical window at $\lambda \approx 100$ m has been partially

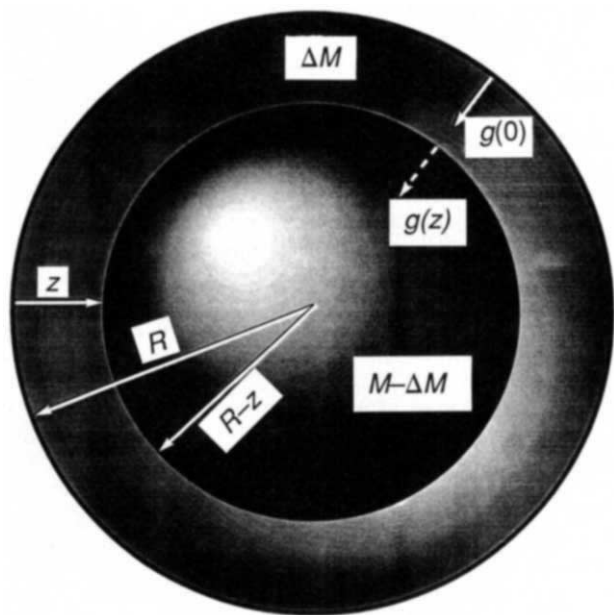


FIG. 2 Sketch of the Airy method for determining the newtonian constant. By comparing the acceleration $g(z)$ at a depth z with the value $g(0)$ at the surface, the shell of mass ΔM can be 'weighed'. See text, ref. 15 and equation (4) for details.

closed by a series of experiments using geophysical sources, as we now discuss. In 1988, Eckhardt and collaborators²⁹ introduced a new technique into the field, by measuring the gravitational acceleration $g(z)$ as a function of height z using the 600-m tower of the television station WTVD in Garner, North Carolina. If newtonian gravity is correct then Laplace's equation can be used to predict $g(z)$, given a knowledge of g on the surface of the Earth near the tower. Although this technique is in principle straightforward, it had never been attempted before 1988, apparently because of the widespread (and erroneous) belief that $g(z)$ could not be measured with sufficient precision on towers. Although measuring $g(z)$ on a tower is indeed more difficult than in the relatively protected environment of a mine-shaft, the tower method has an important compensating advantage. If there are pockets of unusually dense material below the surface of the Earth near the experiment, these will be located on average farther away from the gravimeter being used to determine $g(z)$ in a tower experiment than for analogous measurements in a mine or borehole. The distortion by such local mass concentrations of the average gravitational field near the tower will have less effect than in a similar mine experiment. This 'smoothing' of the average gravitational field facilitates a comparison between the measured and predicted values of $g(z)$. Eckhardt *et al.* in fact demonstrated not only the feasibility of tower measurements, but also the ability of this technique to set useful limits on possible deviations from newtonian gravity.

In their original analysis²⁹, Eckhardt *et al.* found a discrepancy of $(-500 \pm 35) \mu\text{Gal}$ ($1 \mu\text{Gal} = 10^{-8} \text{ m s}^{-2}$) between the measured and predicted values of $g(z)$ at $z = 562.24 \text{ m}$. The negative sign corresponds to the effect expected from a new attractive non-newtonian force, which Eckhardt *et al.* termed the 'sixth' force. But it was subsequently pointed out by Bartlett and Tew³⁰ that the surface gravity database used in ref. 29, which was compiled by the Defense Mapping Agency, could have been biased by oversampling of the higher elevation terrain compared with that at lower elevation. Low-lying terrain is often relatively inaccessible,

as would be the case if a stream were running through the region. As the gravimeters used in such experiments are expensive, it is not surprising that fewer measurements would be made, for instance, at the edge of a body of water, than at a nearby road running above the water. A difference of 1 m between the average elevation of the actual topography and that of the sampled points could lead to a discrepancy of $\sim 309 \mu\text{Gal}$, which is comparable to the magnitude of the effect observed in ref. 29. Eckhardt *et al.* have recently re-examined the effects of the terrain in their experiment, and now find a result consistent with newtonian gravity³¹. Similar null results have also been obtained recently in two other tower experiments^{32,33}. Speake *et al.*³³ devoted considerable effort to understanding how tower motion affects the performance of the LaCoste-Romberg gravimeters used in these experiments. Their careful analysis supported the original claim by Eckhardt *et al.* that $g(z)$ could be measured on towers with sufficient precision to allow a meaningful comparison with the predictions of newtonian gravity. The limits implied by these experiments are partially responsible for filling in the 'geophysical window' in Fig. 1.

Another class of geophysical experiments which tests newtonian gravity over the same distance scales comprises the pumped lake experiments of Müller *et al.*³⁴ and Moore *et al.*³⁵. In these, the change in the local value of g is measured as a function of the height to which the reservoir is filled, and the experiments can thus be viewed as 'weighing' a slab of water whose lateral extent is hundreds of metres. The lake experiments are subject to a number of systematic uncertainties, and thus far have not achieved the sensitivity of the tower experiments.

Similarly, Zumberge *et al.*³⁶ very recently used the ocean as an attracting mass to determine the newtonian constant G for matter interacting over a distance scale of $\sim 5,000 \text{ m}$. This type of Airy experiment, although technically complex, offers a number of advantages over similar measurements in mines, lakes and boreholes. First, ocean experiments allow G to be studied over distance scales that are larger than those readily accessible in land-based experiments. Second, the density of the 5,000-m shell of matter (see Fig. 2) is not only much more uniform than in a mine, but varies in a known and reasonably smooth way with depth. (Mines, after all, are situated where they are precisely because of the inhomogeneities they contain in the form of relatively dense ores.) By combining various sea-surface measurements, Zumberge *et al.* determined G from a generalized version of equation (4). Their result, $G = (6.677 \pm 0.013) \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ agrees with the laboratory value.

This experiment helps fill in a large part of the geophysical window, and sets useful constraints down to values of $\lambda \approx 0.1 \text{ m}$. As noted previously, experiments typically give their most stringent limits for values of λ comparable to the characteristic separation of the test masses in that experiment. It might then seem surprising that an experiment carried out on a distance scale of $\sim 5,000 \text{ m}$ can give a useful limit down to $\sim 0.1 \text{ m}$. But the constraint in Fig. 1 is obtained by comparing the Zumberge value for G to the Cavendish value of G_0 , which is itself determined over a scale of a few centimetres. This discussion applies (with appropriate changes) to the mine/borehole and lake experiments.

The other regimes in which the limits on ξ have improved are at small λ (laboratory scales) and at larger λ (satellite and astronomical scales). At present the best laboratory limits come from the null experiments of Hoskins *et al.*³⁷ and Chen *et al.*³⁸. These experiments arose in part from work by Long³⁹, who in 1976 claimed to see deviations from the inverse-square law. To illustrate the novel technology used in some of these experiments, we will briefly describe the null experiment of Hoskins *et al.*, which tests Newton's law in the 2–5-cm range. These authors note that if the $1/r^2$ law is exact then a test mass located inside an infinitely long hollow cylinder would experience no gravitational force, just as if it were inside a spherical shell. Although there are corrections for end effects in a finite cylinder,

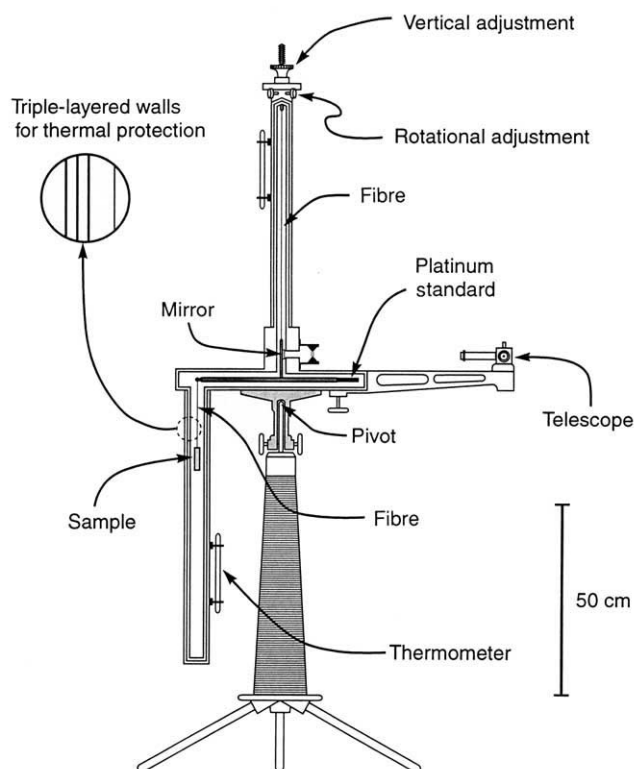


FIG. 3 The Eötvös 'single-arm' apparatus, shown to scale. This drawing is based in part on details supplied in ref. 19, and on an original torsion balance which survives in the museum of the Geophysical Observatory in Tihany, Hungary. See ref. 20 for additional background.

these are small enough to allow detection of the force that would arise from a deviation from the $1/r^2$ law. By monitoring the torque on a test mass inside the cylinder as the distance between the mass and the cylinder is varied, limits can be set on possible non-newtonian forces. Another novel technology is the laplacian detector of Paik^{40,41}, which directly tests the implication of the $1/r^2$ law that the gravitational potential $\Phi(\mathbf{r})$ is a solution of the equation $\nabla^2\Phi(\mathbf{r}) = 0$. Although the existing limits from this experiment are less stringent than those of refs 37 and 38, this technology offers the prospect of considerable improvements in sensitivity. Moreover, the laplacian detector can be used to study newtonian gravity in space, and efforts along this line are currently under way.

For values of λ in the range 10^6 – 10^8 m there has been a considerable improvement in ξ which resulted from a comparison by Rapp⁴² of the values $GM_\oplus$ inferred from Earth-surface gravity measurements, and from a study of the orbits of the LAGEOS satellite and of the Moon about the Earth. The improved limits near $\lambda \approx 10^{11}$ m come from the analysis in ref. 9 of planetary data on $G(r)$, as well as data on anomalous planetary precessions. To understand how these data can be used to set limits on α and λ , we recall from equation (3) that in the presence of a new force the newtonian constant G_∞ is replaced by the function $G(r)$. It follows that under these circumstances Kepler's third law assumes the form

$$a_p^3 = G(a_p)M_\odot(T_p/2\pi)^2 \quad (8)$$

where T_p is the period, and a_p is the physically measured semi-major axis, of planet p . Given a set of values of a_p and T_p for the various planets, $G(a_p)M_\odot$ can be determined and its constancy tested. The curves labelled 'Planetary' in Fig. 1 derive from an even more sensitive test, namely the anomalous precession of the perihelion of the orbit induced by a non-newtonian coupling. For the Yukawa interaction in equation (2) the anomalous precession $\delta\varphi_a$ is given by

$$\delta\varphi_a \approx \pi\alpha(a/\lambda)^2 e^{-a/\lambda} \quad (9)$$

where a is the mean value of the semi-major axis of the orbit. Because a precession of the perihelion is also predicted by general relativity (GR), one can extract a limit on $\delta\varphi_a$ (and thereby on α and λ) only by assuming that GR is correct, for which ample independent evidence exists⁴³.

In summary, much progress has been made since 1981, but additional work needs to be done over the distance scales in the geophysical window. At present Eckhardt and coworkers

are in the process of doing a new experiment using the 2,000-m WABG television tower near Indianola, Mississippi. This site was selected in part because the flatness of the surrounding terrain should help to minimize the problem of terrain bias encountered in their earlier work. Working in the other direction, the terrain problem noted by Bartlett and Tew³⁰ apparently accounts for most of the anomalous results originally reported by Stacey *et al.* But results reported by Ander *et al.*⁴⁴ from Greenland require further study to determine whether the apparent anomalous accelerations $g(z)$ in that experiment could arise from subsurface concentrations of relatively dense material.

Composition-dependent experiments

We turn next to describe the recent progress in composition-dependent experiments. Before 1986, the only reliable data came from the original experiment of Eötvös, Pekár and Fekete (EPF)^{19,20}, which primarily compared the accelerations of various pairs of materials falling to the Earth, and from three later experiments^{26–28} which used the Sun as a source. The Eötvös apparatus (Fig. 3) was used to measure the differential torque on a torsion balance for 11 pairs of chemically different materials¹⁹. As the experiment was originally intended to test the principle of universality of free fall, it was assumed that any acceleration difference between the test samples would arise from the force exerted on these materials from the Earth as a whole. But it was noted in ref. 1 that their results could also be used to set limits on interactions arising from a local source (such as a mountain, cliff, basement), if the chemical composition and matter distribution of the source were sufficiently well known. References 7, 19 and 20 describe the original EPF experiment in more detail.

The original Eötvös experiment can be used to set limits on intermediate-range composition-dependent forces, but because various details of this experiment remain unknown to us, these limits are necessarily somewhat model-dependent. Thus, as a practical matter, the only unambiguous limits on composition-dependent forces before 1986 applied to those for which $\lambda \geq 1$ AU, which derive from the experiments of refs 26–28. We see from Fig. 4 that the whole distance scale from a few centimetres to 1 AU has now been filled in by high-precision experiments using both laboratory and geophysical sources. (As in Fig. 1, we show only the envelope of the limits derived from the various experiments, and not the individual results.) Most of the new experiments use torsion balances similar to those developed by Eötvös to compare the accelerations of their test masses. We

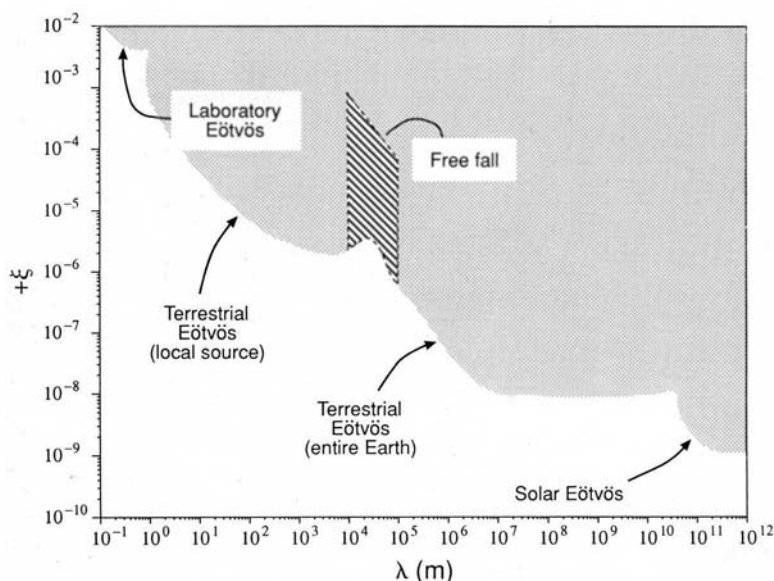


FIG. 4 Constraints on the coupling constant ξ as a function of the range λ from composition-dependent experiments, assuming a coupling to B . With the exception of the region denoted by 'Solar Eötvös', all the constraints shown have been obtained since 1986, and most of the results shown are from the Eöt-Wash experiments⁴⁵. The hatched region denotes values of λ where the limits on ξ depend on detailed models of the Earth and are at present uncertain. The dotted contour is our 'best guess' for the constraint in this region.

show in Fig. 5 the apparatus used Adelberger *et al.*⁴⁵ (known as the Eöt-Wash collaboration), who have carried out the most extensive searches for composition-dependent effects over a wide range of distance scales. As can be seen in Fig. 5, all of the test masses in the Eöt-Wash balance are located in the same vertical position, in contrast to the case for the Eötvös balance. The Eötvös apparatus was originally designed to measure gravity gradients, which required a vertical separation of the masses. This separation is, however, undesirable in searches for composition-dependent effects, as it produces a substantial systematic bias which must be compensated for. The Eöt-Wash balance introduced other innovations, such as compensating masses to cancel (vertical) gravity gradients, a high-precision turntable to rotate the apparatus smoothly and an optical read-out system. Adelberger and collaborators have compared the accelerations of different pairs of test masses towards various sources, including a large mass of lead bricks, the hillside adjacent to their apparatus, and the Earth. In all cases their results have been compatible with the weak equivalence principle, which is to say that they see no evidence for a composition-dependent fifth force. As an example, their most recent hillside limit on the acceleration difference $\Delta a_{\perp}$ of aluminium and beryllium is⁴⁵

$$\Delta a_{\perp}[\text{Be} - \text{Al}] = [(-2.0 \pm 2.2)\hat{\mathbf{E}} + (0.9 \pm 2.1)\hat{\mathbf{N}}] \times 10^{-11} \text{ cm s}^{-2} \quad (10)$$

where $\hat{\mathbf{E}}$ and $\hat{\mathbf{N}}$ are unit vectors which point east and north respectively. For further details of the Eöt-Wash experiments, see ref. 45.

A complementary design for a torsion balance has been introduced by Boynton *et al.*⁴⁶. These authors measure the period $T(\theta)$ of a ring composed of two dissimilar materials oscillating in a horizontal plane, as a function of the initial orientation angle θ of the ring with respect to their source, a cliff located

near Mount Index in Washington. It can be shown⁷ that a dependence of $T(\theta)$ on the orientation of their ring (which is a composition dipole) is a signal for a composition-dependent fifth force, just as the shift in the equilibrium position of the balance is for the EPF and Eöt-Wash experiments. The first results published by Boynton *et al.* found a marginally significant signal that could be interpreted as evidence for a fifth force. Their signal was a nonzero value of the difference $\Delta T(\theta)/T(\theta)$ where $\Delta T(\theta) = T(\theta) - T(\theta + \pi)$, and where $\theta = 0$ corresponded to the aluminium half of their beryllium-aluminium ring being adjacent to the cliff. Boynton *et al.* found for $\Delta T(\theta)$

$$\Delta T(\theta)/T(\theta) = (-4.6 \pm 1.1) \times 10^{-6} \cos \theta + (+0.1 \pm 1.2) \sin \theta \quad (11)$$

But a later experiment⁴⁷ using a copper-polyethylene ring found no evidence for a fifth force. It is not yet clear how we should interpret the result in equation (11), and Boynton *et al.* are continuing their efforts with an improved apparatus. The result in equation (11) is not compatible with the more stringent limits of experiments that obtained null results (see Table 1 of ref. 3).

In addition to variants of torsion balances, a number of novel techniques have also been introduced, such as the floating-ball experiments of Thieberger⁴⁸ and of Bizzeti *et al.*⁴⁹, and the free-fall experiments of Niebauer *et al.*²⁴ and Kuroda and Mio²⁵. Thieberger's experiment is shown schematically in Fig. 6. The central feature is a hollow copper ball floating in water. A composition-dependent fifth-force $\mathbf{F}_5$ (arising in his experiment from a cliff at the Palisades, New Jersey) would act differently on the test masses, which are the copper shell and the water it displaces. This unbalanced force causes the ball to move, and from its velocity we can infer the magnitude of $\mathbf{F}_5$. The advantage of such an apparatus is that the high degree of symmetry of the test masses makes this experiment relatively insensitive to gravity gradients, which could simulate the presence of a fifth-force. In practice, the symmetry of Thieberger's apparatus is broken by a small pin which protrudes from the sphere and which is needed to make the sphere buoyant. Bizzeti *et al.* have carried out a similar experiment (in a monastery near Vallambrosa, Italy) with an apparatus of even higher symmetry. They use a solid nylon sphere floating in a solution containing various salts, and the ball is kept buoyant by a gradient in the density of the solution.

Historically, Thieberger's experiment was the first test of the fifth force hypothesis to search for a composition-dependent effect. In fact Thieberger detected such an effect, in the form of a steady drift of his sphere across the tank. The average velocity of his sphere, $v = (4.7 \pm 0.2) \text{ mm h}^{-1}$, could be accounted for by the model in equations (5)–(7) with $\alpha\lambda = (1.2 \pm 0.4)\text{m}$. By contrast, Bizzeti *et al.* have set an upper limit of $v < 0.010 \text{ mm h}^{-1} (1\sigma)$ on a drift of their sphere, and the corresponding limit for a coupling to B is $\xi\lambda < 0.30 \text{ m} (1\sigma)$ for $\lambda \leq 1 \text{ km}$.

Another interesting route which is being explored is the use of gravity wave antennas as fifth force detectors^{50,51}. In simple conceptual terms, such an experiment consists of a cylindrical detector and a nearby rotating composition dipole whose axis passes through the centre-of-mass of the rotating masses. In the absence of a fifth force the rotor would drive the detector at twice the frequency ω of rotation (assuming that the axis of rotation is precisely aligned with the centre-of-mass of the rotor). In the presence of a composition-dependent fifth force, there would be an additional force at ω , and if ω corresponds to the resonant frequency of the bar, then limits can be set on ξ . At present these limits are not yet competitive with those from other experiments, but as the technology of gravity-wave detectors improves, so will the sensitivity of such experiments.

In addition to composition-dependent experiments that use geophysical sources, other experiments constrain ξ for both smaller and larger values of λ . Using Fig. 4, we can summarize the present situation as follows. For small values of λ , experi-

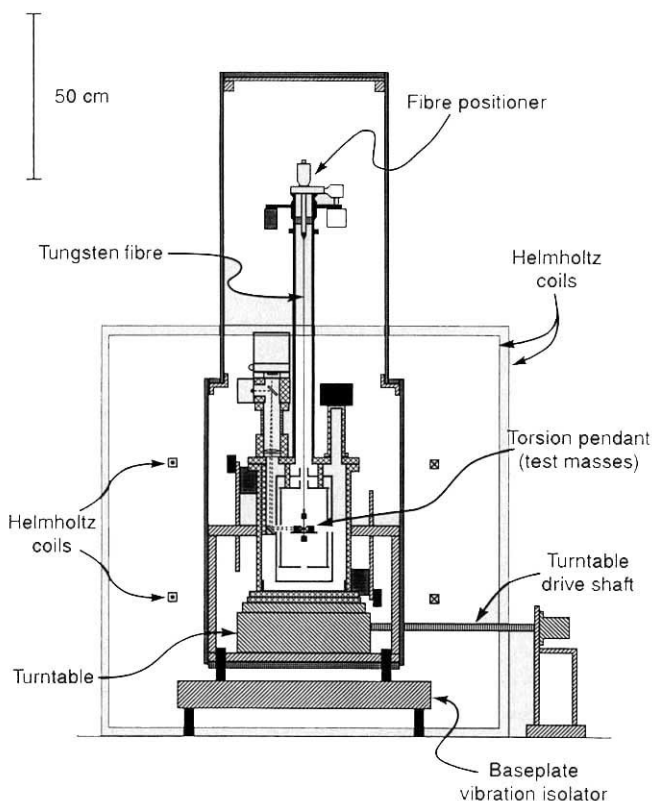


FIG. 5 Diagram of the Eöt-Wash torsion balance of Adelberger *et al.* adapted from ref. 45. In contrast to the original Eötvös apparatus (Fig. 3), the Eöt-Wash balance has the different test masses very symmetrically arranged. The experimental site is the Nuclear Physics Laboratory, located on a hillside at the University of Washington in Seattle.

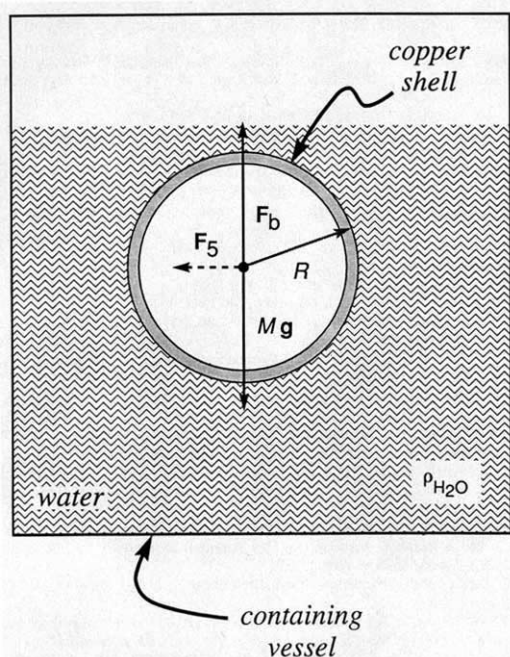


FIG. 6 Schematic drawing of the forces on a hollow copper shell floating in water. In the absence of a composition-dependent fifth force F_5 , the weight Mg is supported by the buoyant force F_b . The presence of F_5 (due to an external source) leads to an unbalanced force which moves the shell. See refs 48 and 49 for more details.

ments have typically involved laboratory-size sources constructed using blocks of lead^{45,52-55}. The obvious advantage of such experiments is that the source has known dimensions and composition, but their disadvantage is that the sensitivity of such experiments drops quickly as the range of the putative fifth force increases. As noted above, for larger λ the best limits on ξ come from experiments using geophysical sources, such as a canal lock⁵⁶, a mountain or a cliff^{45,46,48,49,57}. These have the advantage of using much larger sources than laboratory experiments, albeit ones whose compositions are less well known.

As λ increases, there is a region in which the best limits on ξ come from Galileo-type experiments^{24,25}, which compare the accelerations of two test masses falling freely towards the Earth. These experiments are inherently less sensitive than those using torsion balances, but are in practice easier to analyse in the region $10^4 \text{ m} \leq \lambda \leq 10^6 \text{ m}$. This is a consequence of the fact that Galileo experiments are sensitive to the component of force parallel to $\mathbf{g}$, ($F_{\parallel}$) whereas torsion balance experiments are necessarily sensitive only to the component of $\mathbf{F}$ perpendicular to $\mathbf{g}$. Over this intermediate range, $F_{\parallel}$ is significantly less sensitive to fine details of the underlying structure of the Earth than is $F_{\perp}$, and hence is much more precisely known. As the Galileo experiments can be interpreted with much higher certainty than the torsion balance experiments, they consequently give rise to the best limits in this region of λ . Further improvements in the analysis of $F_{\perp}$ over these distance scales may eventually allow limits from torsion balance measurements to surpass those from Galileo measurements, which are indicated by the hatched region in Fig. 4.

As λ increases still further, terrestrial Eötvös experiments again surpass free-fall experiments in the limits they give for ξ , because $F_{\perp}$ now depends only on global properties of the structure of the Earth. Finally, for $\lambda \approx 1 \text{ AU}$ the Eötvös-type experiments using the Sun as a source are the most sensitive, as we noted earlier. We see that at present, the distance scale from roughly 1 m to 1 AU is now covered by overlapping experiments using a variety of technologies.

When studying composition-dependent effects it is also necessary to take account of the possibility that a given pair of

materials j and j' may coincidentally have the same values of B/μ (ratio of baryon number to m/m_H), and hence have the same accelerations even in the presence of a fifth force. Related to this is the possibility that the charge that determines the strength of the fifth force interaction is not necessarily B , but may be some linear combination of the possible quantum numbers which describe a sample of bulk matter. If N and Z denote the numbers of neutrons and protons respectively, then the most general additive charge Q is

$$Q = \cos \theta_5 B + \sin \theta_5 I_Z \quad (12)$$

where (as before) $B = N + Z$, $I_Z = N - Z$ and θ_5 is an appropriate 'mixing angle'. The previous formalism can then be taken over with B/μ replaced by Q/μ . It follows from equation (12) that at least two pairs of materials must be compared to rule out the possibility of an accidental cancellation in the charges of the test masses, and this has been checked explicitly by the Eöt-Wash collaboration⁴⁵. It is also possible that for some choices of Q (for example $Q \equiv I_Z$) the source strength $(Q/\mu)_i$ could also vanish. For $Q = I_Z$ this is an interesting possibility, because most geophysical sources have near-zero values of I_Z . (An exception is water, which was the source in the experiment of Bennett⁵⁶.) The possibility that some geophysical experiments had obtained null results because $Q = I_Z \equiv 0$ was excluded by various laboratory experiments^{45,52-55} which capitalized on the fact that a lead source has a relatively large value of I_Z .

Elementary particle experiments

The energy-dependent effects reported by Aronson *et al.*¹⁶ which provided part of the motivation for the fifth-force hypothesis, have also been tested experimentally. Grossman *et al.*⁵⁸ searched for an energy dependence of the K_S lifetime $\tau_S = \hbar/\Gamma_S$ by comparing its value measured in the momentum range (100–350) GeV/c to the laboratory value determined in the range (2–5) GeV/c. Their result, $\tau_S = (0.8920 \pm 0.0044) \times 10^{-10} \text{ s}$, is in excellent agreement with the laboratory value $\tau_S = (0.8922 \pm 0.0020) \times 10^{-10} \text{ s}$. Similarly, Coupal *et al.*⁵⁹ found $|\eta_{\pm}| = (2.28 \pm 0.06) \times 10^{-3}$ in the momentum range (30–150) GeV/c, which agrees well with the low-energy laboratory value $|\eta_{\pm}| = (2.268 \pm 0.023) \times 10^{-3}$. But as noted in ref. 60 the Coupal result for $|\eta_{\pm}|$ is also consistent, within their quoted errors, with that of Aronson *et al.*¹⁶. Recently, Carosi *et al.*⁶¹ have reported high-energy determinations of $\phi_{\pm}$ and ϕ_{00} . They obtain $\phi_{\pm} = 46.9^\circ \pm 1.4^\circ \pm 0.7^\circ$, $\phi_{00} = 47.1^\circ \pm 2.1^\circ \pm 1.0^\circ$, $70 \text{ GeV} \leq E_K \leq 170 \text{ GeV}$, where the first error is statistical and the second is systematic. These results are in good agreement with the corresponding low-energy values⁶².

Another implication of the original model of a coupling to hypercharge has also been called into question by recent experiments. This is the suggestion⁶³⁻⁶⁶ that the quanta of this field, called hyperphotons (γ_Y), could be detected in decays such as $K^+ \rightarrow \pi^+ + \gamma_Y$. The signal for such a decay would be the detection of a π^+ with an appropriate momentum, $|\mathbf{p}_{\pi}| = 227 \text{ MeV}/c$, and no other detected particles. This is a fairly clean signal which has been looked for in several recent experiments. The current best limits on this branching ratio are

$$\frac{\Gamma(K^+ \rightarrow \pi^+ + \text{undetected neutrals})}{\Gamma(K^+ \rightarrow \text{all final states})} \leq \begin{cases} 4.6 \times 10^{-8}, & 90\% \text{ confidence limits}^{67} \\ 6.4 \times 10^{-9}, & 90\% \text{ confidence limits}^{68} \end{cases} \quad (13)$$

Several theoretical calculations have used the experimental limit in equation (13) above to infer limits on the strength, f^2 , of a possible coupling to hypercharge. Although the results of these calculations differ considerably, for a coupling to hypercharge ($Y = B + S$) they all agree that the inferred values of f^2 are probably too small to be compatible with either the energy-dependence of the kaon parameters claimed by Aronson *et al.* or with the value of f^2 required to explain the Eötvös data¹⁹.

We emphasize that this conclusion applies specifically to the case where γ_V is a vector particle and couples to hypercharge; for a scalar field there are no meaningful limits from kaon decay. Similarly, if γ_V couples to some more general linear combination of B and S , the constraints arising from K^+ decay are not in direct conflict with other data (see ref. 60 for more details).

Another class of elementary-particle experiments that can be used to explore models of the putative fifth force consists of those comparing the free-fall accelerations of particles and their antiparticles. In the 1960s Fairbank and coworkers⁶⁹ attempted to measure the gravitational accelerations of e^+ and e^- , but these efforts were thwarted by the presence of residual electromagnetic fields produced by surface effects in their apparatus. Goldman, Hughes and Nieto (GHN)⁷⁰ subsequently noted that such effects would be relatively less important in an experiment comparing the accelerations of p and $\bar{p}$, because of the larger masses of these particles. The impetus for carrying out such an experiment increased following the suggestion of a possible fifth force, as the proposed (vector) hypercharge field would be expected to produce a difference in the apparent gravitational accelerations of p and $\bar{p}$. In most models this difference would be too small to be detected, but GHN also noted that some models containing both scalar and vector fields predicted an effect large enough to be measurable.

This claim has been called into question by several recent analyses^{71–74}. As was first noted by Schiff⁷⁵, if matter and antimatter behaved differently in a gravitational field, this could show up as an anomaly in the Eötvös experiment. Using a similar argument, Adelberger *et al.*⁷¹ demonstrate that a fifth-force model that could produce an acceleration difference between p and $\bar{p}$ would also produce an acceleration difference between the test masses in a typical Eötvös experiment. Given the extraordinary sensitivity of the current Eötvös experiments, Adelberger *et al.* argue that existing limits from their Eöt-Wash experiments preclude seeing any acceleration difference between p and $\bar{p}$ at the sensitivity level expected in these experiments. See refs 71–74 and 76 for further discussion.

Pushing the limits

The experimental searches for both composition-independent and composition-dependent deviations from newtonian gravity have made enormous advances since 1986, as shown in Figs 1 and 4. No compelling evidence has yet emerged that would indicate the presence of a fifth force, although the anomalies reported in the original Eötvös experiment remain to be understood, as do those in the experiments of Thieberger⁴⁸ and Boynton *et al.*⁴⁶. On the experimental side, efforts continue to set even more stringent limits on possible deviations from newtonian gravity, motivated in part by the recognition that such experiments may be our most powerful tool in exploring physics at the Planck scale. The connection, to which we have alluded earlier, between such experiments and physics at the Planck scale has been explored in several models^{70,77–81}. An excellent review of the wide class of theories that predict the existence of weak macroscopic forces has been given recently by Fujii⁸¹. Although it would be difficult to summarize the content of these models briefly, the main lesson is simply that there are many theories that predict such forces. It follows that if additional weak forces are not seen, as is the indication from present experiments, then the assumptions behind a wide class of theories will have been called into question.

The experiments we have discussed can be understood as extending the maximum energy scale that can be explored from 4×10^4 GeV (at the proposed Superconducting Super-Collider), to $\sim 10^{19}$ GeV, an energy beyond the capability of any accelerator. Viewed in this way, torsion balances, floating balls and tall towers may be the ultimate high-energy physics tools. $\square$

Ephraim Fischbach and Carrick Talmadge are at the Physics Department, Purdue University, West Lafayette, Indiana 47907, USA.

- Fischbach, E., Sudarsky, D., Szafer, A., Talmadge, C. & Aronson, S. H. *Phys. Rev. Lett.* **56**, 3–6 (1986).
- Fischbach, E., Gillies, G. T., Krause, D. E., Schwan, J. G. & Talmadge, C. *Metrologia* (in the press).
- Adelberger, E. G., Heckel, B. R., Stubbs, C. W. & Rogers, W. F. A. *Rev. nucl. Part. Sci.* **41**, 269–320 (1991).
- Fujii, Y. *Nature* **234**, 5–7 (1971); *Ann. Phys.* **69**, 494–521 (1971).
- Wagoner, R. V. *Phys. Rev. D* **1**, 3209–3216 (1970).
- O'Hanlon, J. *Phys. Rev. Lett.* **29**, 137–138 (1972).
- Fischbach, E., Sudarsky, D., Szafer, A., Talmadge, C. & Aronson, S. H. *Ann. Phys.* **182**, 1–89 (1988).
- Gibbons, G. W. & Whiting, B. F. *Nature* **291**, 636–638 (1981).
- Talmadge, C., Berthias, J.-P., Hellings, R. W. & Standish, E. M. *Phys. Rev. Lett.* **61**, 1159–1162 (1988).
- De Rújula, A. *Phys. Lett.* **B180**, 213–220 (1986).
- Stacey, F. D. & Tuck, G. J. *Nature* **292**, 230–232 (1981).
- Holding, S. C. & Tuck, G. J. *Nature* **307**, 714–716 (1984).
- Stacey, F. D. *et al. Rev. Mod. Phys.* **59**, 157–174 (1987).
- Stacey, F. D., Tuck, G. J. & Moore, G. I. *Phys. Rev. D* **36**, 2374–2380 (1987).
- Stacey, F. D. *Sci. Prog.* **69**, 1–17 (1984).
- Aronson, S. H., Bock, G. J., Cheng, H.-Y. & Fischbach, E. *Phys. Rev. Lett.* **48**, 1306–1309 (1982); *Phys. Rev. D* **28**, 476–494 (1983); *Phys. Rev. D* **28**, 495–523 (1983).
- Sudarsky, D., Fischbach, E., Talmadge, C., Aronson, S. H. & Cheng, H.-Y. *Ann. Phys.* **207**, 103–139 (1991).
- Fischbach, E., Cheng, H.-Y., Aronson, S. H. & Bock, G. J. *Phys. Lett.* **B116**, 73–76 (1982).
- Eötvös, R. v., Pekár, D. & Fekete, E. *Ann. Phys. Leipzig* **68**, 11–66 (1922); *Roland Eötvös Gesamte Werke* (ed. Selényi, P.) 307–372 (Akadémiai Kiadó, Budapest, 1953); (Engl. transl.) Achzehtner, J. *et al. Univ. Washington Preprint* 40048-13-N6; *Ann. Univ. Sci. Budap. Rolando Eötvös Nominata, Sect. Geol.* **7**, 111–165 (1963).
- Bod, L., Fischbach, E., Marx, G. & Náráy-Ziegler, M. *Acta phys. Hungarica* **69**, 335–355 (1991).
- Galileo Galilei *Dialogues Concerning Two New Sciences*, 1638 edn (transl. Crew, H. & de Salvio, A.) 212–213 (Macmillan, New York, 1914).
- Mach, E. *The Science of Mechanics* (transl. McCormack, T. J.) 151 (Open Court, La Salle Illinois, 1960).
- Misner, C. W., Thorne, K. S. & Wheeler, J. A. *Gravitation* 16 (Freeman, San Francisco, 1973).
- Niebauer, T. M., McHugh, M. P. & Faller, J. E. *Phys. Rev. Lett.* **59**, 609–612 (1987).
- Kuroda, K. & Mio, N. *Phys. Rev. Lett.* **62**, 1941–1944 (1989); *Phys. Rev. D* **42**, 3903–3907 (1990).
- Roll, P. G., Krotkov, R. & Dicke, R. H. *Ann. Phys.* **26**, 442–517 (1964).
- Braginskii, V. B. & Panov, V. I. *Soviet Phys. JETP* **34**, 463–466 (1972).
- Keiser, G. M. & Faller, J. E. in *Proc. 2nd Marcel Grossmann Meet. General Relativity* (ed. Ruffini, R.) 969–976 (North-Holland, Amsterdam, 1982).
- Eckhardt, D. H., Jekeli, C., Lazarewicz, A. R., Romaides, A. J. & Sands, R. W. *Phys. Rev. Lett.* **60**, 2567–2570 (1988).
- Bartlett, D. F. & Tew, W. L. *Phys. Rev. Lett.* **63**, 1531 (1989); *J. Geophys. Res.* **95**, 17,363–17,369 (1990).
- Jekeli, C., Eckhardt, D. H. & Romaides, A. J. *Phys. Rev. Lett.* **64**, 1204–1206 (1990).
- Thomas, J. *et al. Phys. Rev. Lett.* **63**, 1902–1905 (1989).
- Speake, C. C. *et al. Phys. Rev. Lett.* **65**, 1967–1971 (1990).
- Müller, G., Zürn, W., Lindner, K. & Rösch, N. *Phys. Rev. Lett.* **63**, 2621–2624 (1989); *Geophys. J. Int.* **101**, 329–344 (1990).
- Moore, G. I., Stacey, F. D., Tuck, G. J., Goodwin, B. D. & Roth, A. J. *phys. Eng. Sci. Instr.* **21**, 534–539 (1988).
- Zumberge, M. A. *et al. Phys. Rev. Lett.* **67**, 3051–3054 (1991).
- Hoskins, J. S., Newman, R. D., Spero, R. & Schultz, J. *Phys. Rev. D* **32**, 3084–3095 (1985).
- Chen, Y. T., Cook, A. H. & Metherell, A. J. F. *Proc. R. Soc. A* **394**, 47–68 (1984).
- Long, D. *Nature* **260**, 417–418 (1976).
- Paik, H. J. *Phys. Rev. D* **19**, 2320–2324 (1979).
- Chan, H. A., Moody, M. V. & Paik, H. J. *Phys. Rev. Lett.* **49**, 1745–1748 (1982).
- Rapp, R. H. *Geophys. Res. Lett.* **14**, 730–732 (1987).
- Will, C. M. *Science* **250**, 770–776 (1990); *Phys. Rep.* **113**, 345–422 (1984); *Theory and Experiment in Gravitational Physics* (Cambridge Univ. Press, 1981).
- Ander, M. E. *et al. Phys. Rev. Lett.* **62**, 985–988 (1989).
- Adelberger, E. G. *et al. Phys. Rev. D* **42**, 3267–3292 (1990).
- Boynton, P. E., Crosby, D., Ekstrom, P. & Szumilo, A. *Phys. Rev. Lett.* **59**, 1385–1389 (1987).
- Boynton, P. E. & Aronson, S. H. in *New and Exotic Phenomena '90* (eds Fackler, O. & Trần Thanh Vân, J.) 207–224 (Editions Frontières, Gif-sur-Yvette, 1990).
- Thieberger, P. *Phys. Rev. Lett.* **60**, 965 (1988).
- Bizzetti, P. G., Bizzetti-Sona, A. M., Fazzini, T., Perego, A. & Taccetti, N. *Phys. Rev. Lett.* **62**, 2901–2904 (1989); in *New and Exotic Phenomena '90* (eds Fackler, O. & Trần Thanh Vân, J.) 263–268 (Editions Frontières, Gif-sur-Yvette, 1990).
- Akasaka, N., Hirakawa, H., Mio, N., Ohashi, M. & Tsubono, K. in *Proc. 5th Marcel Grossmann Meet. General Relativity* (eds Blair, D. G. & Buckingham, M. J.) 1591–1594 (World Scientific, Singapore, 1989).
- Astone, F. *et al. Z. Phys.* **C50**, 21–29 (1991).
- Nelson, P., Graham, D. M. & Newman, R. D. *Phys. Rev. D* **42**, 963–976 (1990).
- Speake, C. C. & Quinn, T. J. *Phys. Rev. Lett.* **61**, 1340–1343 (1988).
- Stubbs, C. W. *et al. Phys. Rev. Lett.* **62**, 609–612 (1989).
- Cowsik, R., Krishnan, N., Tandon, S. N. & Unnikrishnan, S. *Phys. Rev. Lett.* **61**, 2179–2181 (1988); *Phys. Rev. Lett.* **64**, 336–339 (1990).
- Bennett, W. R. Jr *Phys. Rev. Lett.* **62**, 365–368 (1989).
- Fitch, V., Isaila, M. V. & Palmer, M. A. *Phys. Rev. Lett.* **60**, 1801–1804 (1988).
- Grossman, N. *et al. Phys. Rev. Lett.* **59**, 18–21 (1987).
- Coupal, D. P. *et al. Phys. Rev. Lett.* **55**, 566–569 (1985).
- Aronson, S. H., Fischbach, E., Sudarsky, D. & Talmadge, C. in *5th Force: Neutrino Physics* (eds Fackler, O. & Trần Thanh Vân, J.) 593–602 (Editions Frontières, Gif-sur-Yvette, 1988).
- Carosi, R. *et al. Phys. Lett. B* **237**, 303–312 (1990).
- Carosi, R. *et al. Phys. Lett. B* **239**, 1–516 (1990).
- Suzuki, M. *Phys. Rev. Lett.* **56**, 1339–1341 (1986).
- Aronson, S. H., Cheng, H.-Y., Fischbach, E. & Haxton, W. *Phys. Rev. Lett.* **56**, 1342–1345 (1986).
- Bouchiat, C. & Iliopoulos, J. *Phys. Lett. B* **169**, 447–449 (1986).
- Lisignoli, M. & Pugliese, A. *Phys. Lett. B* **171**, 468–470 (1986).
- Asano, Y. *et al. Phys. Lett.* **107**, 159–162 (1981).
- Littenberg, L. in *Proc. 1989 Int. Symp. Lepton and Photon Interactions at High Energies August 7–12, 1989* (ed. Riordan, M.) 184–201 (World Scientific, Singapore, 1989).
- Fairbank, W. M., Witteborn, F. C., Madey, J. M. J. & Lockhart, J. M. in *Experimental Gravitation Lake Como, Italy, July 17–29 1972* (ed. Bertotti, B.) 310–330 (Academic, New York, 1974).
- Goldman, T., Hughes, R. J. & Nieto, M. M. *Phys. Lett. B* **171**, 217–222 (1986); *Scient. Am.* **258**, 48–56 (1988).
- Adelberger, E. G., Heckel, B. R., Stubbs, C. W. & Su, Y. *Phys. Rev. Lett.* **66**, 850–853 (1991).
- Morgugo, G. *Phys. Rev. Lett.* **67**, 1047 (1991).

73. Goldman, T. *et al.* *Phys. Rev. Lett.* **67**, 1048 (1991).
 74. Adelberger, E. G. & Heckel, B. R. *Phys. Rev. Lett.* **67**, 1049 (1991).
 75. Schiff, L. I. *Proc. natn. Acad. Sci. U.S.A.* **45**, 69–80 (1959).
 76. Nieto, M. M. & Goldman, T. *Phys. Rep.* **205**, 221–281 (1991).
 77. Peccei, R. D., Sola, J. & Wetterich, C. *Phys. Lett.* **B195**, 183–190 (1987).
 78. Barr, S. M. & Mohapatra, R. N. *Phys. Rev. Lett.* **57**, 3129–3132 (1986).
 79. Halprin, A., Barnhill, M. V. III & Barr, S. M. *Phys. Rev.* **D39**, 1467–1470 (1989).

80. Hill, C. T. & Ross, G. G. *Nucl. Phys.* **B311**, 253–297 (1988).
 81. Fujii, Y. *Int. J. mod. Phys.* **A6**, 3505–3557 (1991).
 82. Scherk, J. *Phys. Lett.* **B88**, 265–267 (1979).

ACKNOWLEDGEMENTS. We thank our many colleagues for discussions which helped to clarify topics covered here. This work was supported by the US Department of Energy and the Air Force Geophysics Laboratory.

ARTICLES

Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours

Lawrence A. Donehower*, Michele Harvey*, Betty L. Slagle*, Mark J. McArthur†, Charles A. Montgomery Jr†, Janet S. Butel* & Allan Bradley‡

* Division of Molecular Virology, † Center for Comparative Medicine, and ‡ Institute for Molecular Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

Mutations in the p53 tumour-suppressor gene are the most frequently observed genetic lesions in human cancers. To investigate the role of the p53 gene in mammalian development and tumorigenesis, a null mutation was introduced into the gene by homologous recombination in murine embryonic stem cells. Mice homozygous for the null allele appear normal but are prone to the spontaneous development of a variety of neoplasms by 6 months of age. These observations indicate that a normal p53 gene is dispensable for embryonic development, that its absence predisposes the animal to neoplastic disease, and that an oncogenic mutant form of p53 is not obligatory for the genesis of many types of tumours.

THE cellular p53 protein was first detected in rodent cells transformed by simian virus SV40 in a complex with the SV40 large T-antigen^{1,2}. Subsequently p53 was shown to be complexed with adenovirus and oncogenic papillomavirus oncoproteins^{3,4}. Initially considered to be a cellular proto-oncogene, recent observations indicate that the gene encoding p53 is a tumour-suppressor gene in its native form but that it may behave as an oncogene in some mutant forms. These mutant forms can immortalize primary cells and cooperate with *ras* to transform primary rodent cells^{5–8}, whereas wild-type p53 actively inhibits transformation by mutant p53 gene plus *ras*^{9,10}. Human tumour cells transfected with a wild-type p53 gene grow at reduced rates in culture^{11–14} and have decreased tumorigenicity¹⁵, indicating a suppression of the tumour phenotype. Deletions, rearrangements and retrovirus insertional inactivations of the p53 gene have been noted in mouse erythroleukaemias^{16,17}, suggesting that loss of normal p53 function may contribute to the emergence of the malignant phenotype.

Mutations and allele loss of the p53 gene have been associated with tumours from a wide variety of human organs (including lung, breast, colon, oesophagus, liver, bladder, ovary and brain) and haematopoietic tissue¹⁸. Mutations in the p53 gene are now the most frequently observed genetic lesion in spontaneous human cancers¹⁹. In addition, p53 mutations are associated with the inherited cancer-susceptibility Li-Fraumeni syndrome^{20,21}.

Affected individuals often inherit a point mutation in highly conserved regions of the p53 gene, primarily in exon 7. Presumably these mutations predispose affected individuals to cancer development, particularly after the normal p53 allele is somatically lost or mutated.

The normal physiological role of p53 seems to be in regulation of the cell cycle. The p53 protein is phosphorylated in a cell-cycle-dependent pattern by cdc2 kinase^{22,23}, a regulatory kinase required for mitosis in mammalian cells, with maximal levels of p53 phosphorylation reached during mitosis²³. In the cell p53 is spatially regulated, accumulating in the cytoplasm during G1 and migrating to the nucleus at the beginning of the S phase²⁴. Transformed cells transfected with a wild-type p53 gene arrest in the G1 phase of the cell cycle^{11–13}. There are two hypotheses^{25,26} to explain the mechanism of action of p53 in regulation of cell growth. One proposes that p53 regulates the initiation of DNA replication. Wild-type murine p53 blocks the interaction of DNA polymerase α with SV40 large T-antigen and prevents SV40 DNA replication^{27,28}, and it may bind to an analogous cellular replication protein to prevent entry into S phase. The second proposes that p53 regulates cell growth by acting as a transactivator of transcription (directly or indirectly), either by stimulation of growth inhibitory genes or by repression of growth stimulatory genes. In yeast, wild-type p53 fusion proteins have transactivating activity^{29,30}.

As p53 is implicated in cell-cycle control, it might be essential for embryonic development. Although p53 is expressed in many tissues in the mouse, total protein levels are low owing to its relatively short half-life, which averages about 20 min. The amount of p53 messenger RNA in developing mouse embryos is maximal at 9–11 days gestation³¹, after which levels fall markedly.

To generate an animal model to investigate the role of p53 in mammalian development and in tumour formation, we used homologous recombination in mouse embryonic stem cells to derive a null allele of the gene. Chimaeric mice were generated, and the mutated p53 allele was established in the mouse germ line. Mice homozygous for the mutated gene are viable and appear completely normal but are susceptible to the spontaneous development of different types of tumours before they reach 6 months of age, particularly lymphomas and sarcomas. Heterozygote progeny carrying a single wild-type p53 allele rarely develop spontaneous tumours before they are 9 months old. These observations indicate that a normal p53 gene is not required for mouse development. In addition, the loss of a normal p53 allele is clearly sufficient to predispose an animal to many types of

tumours; an oncogenic mutant form is not necessary for tumorigenesis.

p53 gene targeting

Our gene targeting protocol was a modified version of a positive-negative selection strategy³². The targeting vector contained 3.7 kilobases (kb) of the genomic p53 gene interrupted in exon 5 by a *PolII-neo* expression cassette³³ that was missing a polyadenylation signal (to provide an enrichment for homologous recombination events) (Fig. 1). The insertion of the *neo* cassette was also accompanied by deletion of a 450-base-pair p53 gene fragment containing 106 nucleotides of exon 5 (including the highly conserved p53 domain II³⁴) and about 350 nucleotides of intron 4.

The p53 targeting construct was electroporated into 2×10^8 AB1 (ref. 35) embryonic stem cells. Simultaneous selection in G418 and FIAU yielded ~100 G418-, FIAU-resistant colonies. Clones were screened by polymerase chain reaction (PCR) for a gene-targeting-specific junction fragment (Fig. 1).

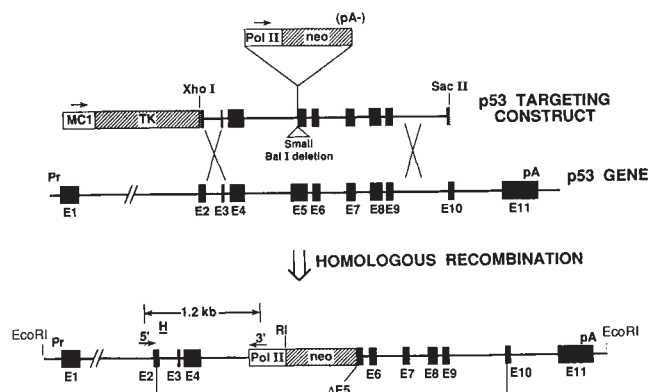


FIG. 1 Disruption of the p53 gene in mouse embryonic stem (ES) cells. The targeting construct used to disrupt the endogenous p53 gene in ES cells is shown at the top. The hypothetical crossovers between the endogenous p53 allele and the targeting construct are indicated by the crossed lines. These crossovers will integrate the poly(A)-deficient (PA⁻) *neo* cassette into the p53 gene and generate a fusion RNA polyadenylated at the p53 poly(A) site. Control experiments indicated that removal of the *neo* poly(A) site provided an enrichment factor of 50-fold for targeted events. Control experiments also verified that the enrichment factor provided by the herpes simplex thymidine kinase (TK) gene was 3-fold. These selections were very specific for crossovers on both the 5' and 3' side of the *neo* insertion. The desired disrupted p53 allele is shown at the bottom. The boundaries of the targeting construct are indicated by vertical lines below the altered gene. The arrows labelled 5' and 3' represent PCR primers used to identify homologous recombination events in ES cell colonies electroporated with the targeting construct. The first primer (5'-TTTACGGAGCCCTGGCGCTC-GATGT-3') is in the *PolII* promoter region and the second primer (5'-ATGACTGCCATGGAGGAGTCACAGTC-3') is in exon 2 of the p53 gene and is not contained in the targeting vector. These primers would generate a 1.2-kb fragment in successfully targeted cells which was then hybridized with the internal oligonucleotide labelled 'H' (5'-GGATATCAGCCTCGAGCTCCCTCT-3'), directly adjacent to the second primer in p53 exon 2 (E2), to verify that the fragment was the correct one. PCR was done as described³⁵, except that clones were treated individually and were not pooled. Two positive clones were grown up and analysed further. Abbreviations: RI, *EcoRI* restriction site; Pr, endogenous p53 promoter; MC1, synthetic promoter for expression in ES cells (Stratagene).

METHODS. The p53 targeting vector was constructed by insertion of the 3.7-kb *XhoI* (exon 2)–*SacII* (exon 10) murine genomic p53 DNA fragment⁴⁴ into pBluescript II KS (Stratagene) to generate pKS-p53. A *PolII-neo* expression cassette³³ (without a polyadenylation signal) was then ligated to the *BalI* sites of pKS-p53 in the same transcriptional orientation as the p53 gene. As there are two *BalI* sites in pKS-p53 (in intron 4 and exon 5), the ligation reaction resulted in a 450-bp deletion of p53, including 106 bp of exon 5. An MC1-TK cassette was then inserted at the 5' end of the p53 gene at the *XhoI* site in the same transcriptional orientation as the p53 gene. The resulting targeting construct was amplified and purified before introduction into ES cells.

Two colonies were positive for the diagnostic 1.2-kb PCR fragment. The structure of the mutant p53 allele was confirmed to have undergone the gene replacement by Southern blot hybridization using three probes: a *neo* probe, a p53 exon 1–2 probe, and a p53 exon 5–10 probe (Fig. 2). The p53 gene is flanked by two *EcoRI* sites about 16 kb apart. The *neo* cassette contains an *EcoRI* site. Thus, as a result of the targeted integration, a new *EcoRI* site is inserted midway between the endogenous sites to generate two novel *EcoRI* fragments of about 8.5 kb. Hybridization with the three probes showed that the predicted 8.5-kb fragments were in genomic DNA in both PCR-positive clones, indicating the correct targeting in the p53 locus. Southern hybridization experiments using restriction enzymes other than *EcoRI* confirmed that the desired recombination event had taken place in the two embryonic stem cell clones (data not shown).

p53-deficient mice

One of the embryonic stem cell clones with a targeted p53 allele was used to generate chimaeric mice by microinjection into 3.5-day-old C57BL/6 blastocysts and implantation of the blastocysts into pseudopregnant F1 (C57BL×CBA) females³⁶. Many of the resulting offspring were chimaeric, as indicated by the predominance of embryonic stem cell-derived (129/Sv strain) Agouti pigmentation in the coat. The chimaeric males were mated to C57BL/6 females. Four chimaeric males were fertile and three were germ-line chimaeras, as indicated by transmission of the Agouti coat colour to their progeny. All three germ-line chimaeras transmitted mutant p53 alleles to their Agouti offspring at the expected 50% frequency.

To assess the role of the p53 gene in development, we intercrossed p53 heterozygote animals (carrying one wild-type p53 allele and one null allele). One-fourth of the progeny from such crosses should inherit the mutant p53 allele from each parent

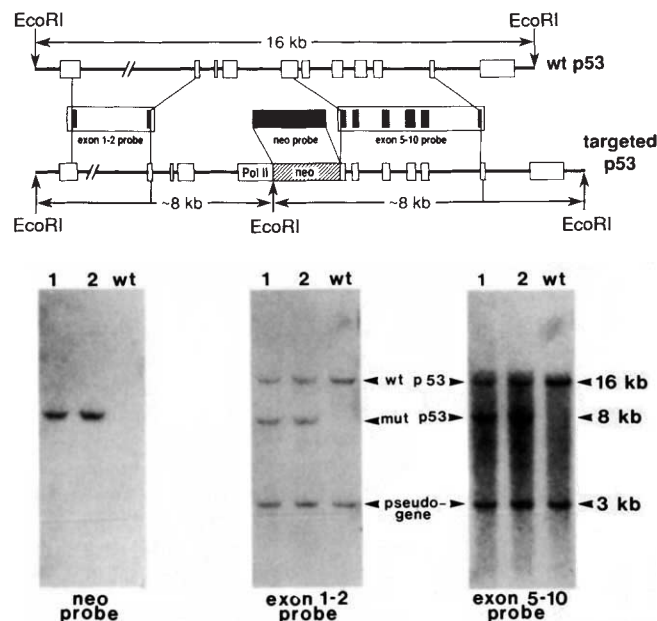


FIG. 2 Southern blot hybridization analysis of ES cell clones with putative homologous recombination events. The two ES cell clones that were positive by PCR analysis for homologous recombination with the p53 targeting construct were analysed by Southern hybridization. ES cell DNA (10 µg) was restricted with *EcoRI*, electrophoresed on a 0.7% agarose gel, blotted onto nylon, and hybridized sequentially to three different ³²P-labelled probes. The derivation of each of the probes is indicated in the diagram showing the structure of the normal and disrupted p53 alleles at the top. Note the added internal *EcoRI* site in the disrupted allele. Hybridization results with each probe are shown in the bottom panels. The hybridization patterns conformed to those expected for a correct homologous recombination event in the two ES cell clone DNAs. A control DNA (labelled 'wt') derived from normal ES cells is shown as the right lane on each blot.

and be homozygous for a disrupted p53 gene (referred to hereafter as 'homozygotes'). Litter sizes were normal, and analysis of tail biopsies at 4 weeks of age from 131 offspring from heterozygote crosses revealed the presence of homozygous mutant mice at a frequency of 23% (Fig. 3). The homozygotes appeared to be normal, both morphologically and by histopathological analyses.

Confirmation of a p53 null allele

Because viable homozygote animals were not anticipated, we next established that the targeted p53 allele was indeed a null allele and that normal p53 protein was not synthesized in tissues of the homozygotes. Our gene replacement vector should have deleted 450 base pairs (bp) of the p53 gene, including the exon 5 splice acceptor site and 106 bp of highly conserved exon 5 sequences. A 106-bp probe specific for the exon 5 sequences deleted in the targeting construct was hybridized to genomic DNAs derived from the offspring of the heterozygote cross shown in Fig. 3 (middle). The homozygotes did not contain any DNA fragments that hybridized to the exon 5 probe (other than

the pseudogene) (Fig. 3, bottom), confirming that both copies of a highly conserved segment of the p53 gene were deleted.

The deletion of the exon 5 splice acceptor site in the targeted allele theoretically could result in cryptic splicing or alternative splicing events. To determine whether intact or quasi-intact p53 mRNAs were generated in tissues of the homozygotes, we analysed total RNA derived from mouse embryo fibroblasts by northern blotting. The homozygote fibroblasts were negative for detectable p53 mRNA of any size, whereas the heterozygote and wild-type cells contained the expected 2.0-kb p53 mRNA (data not shown). To establish whether p53 message was being produced at low levels in the homozygote cells, we employed an RNA-PCR strategy. Five sets of primers were used to amplify complementary DNA synthesized from total RNA purified from spleen, kidney and testes of wild-type, heterozygote and

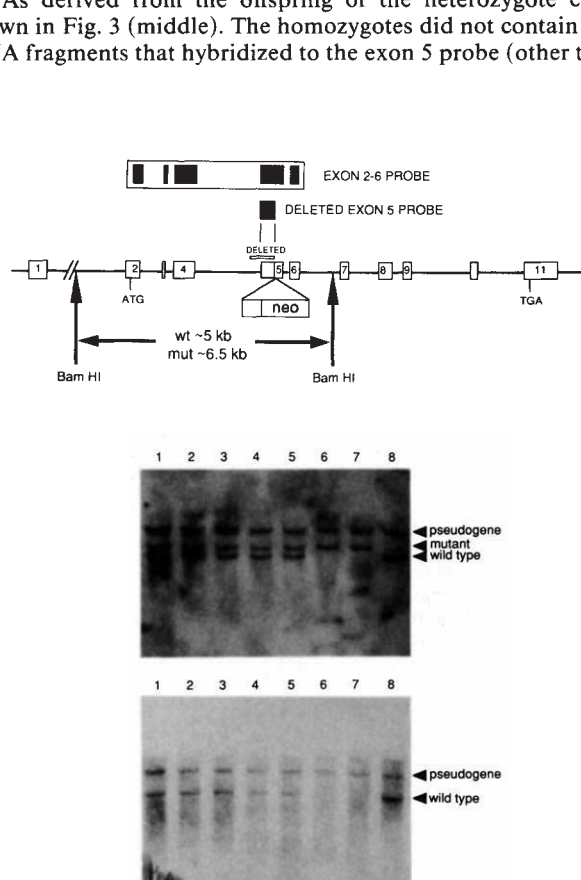


FIG. 3 Mice homozygous for the disrupted p53 allele are viable and lack exon 5 sequences. Genomic DNA (8 μ g) was isolated from the tails of a litter of eight pups derived from a cross of two heterozygous mice. The DNAs were cleaved with *Bam*HI, subjected to agarose gel electrophoresis, blotted onto nylon, and hybridized to one of the two probes shown in the top diagram. The sizes of the expected wild-type and disrupted p53 fragments following *Bam*HI cleavage are indicated. The middle panel shows the results of hybridization of the litter DNAs with an exon 2-6 probe. Lanes 1-5 represent mice heterozygous for the disrupted p53 allele, lanes 6 and 7 show mice homozygous for the disrupted allele, and lane 8 represents a wild-type genotype. The bottom panel shows a Southern blot of the same mouse DNAs as in the middle panel, but hybridized with a probe obtained from the portion of exon 5 deleted in the targeting construct. The 106-bp deletion-specific probe was generated by PCR on a p53 plasmid template using exon 5 primers derived from the boundaries of the exon 5 deletion (5' primer: 5'-TACTCTCCTCCCTCAATAA-3'; 3' primer: 5'-CCATGGCGG-GACACGGCTCC-3'). Note that homozygous mice (lanes 6 and 7) contain no DNA sequences (other than the pseudogene) that hybridize to the p53 exon 5 probe.

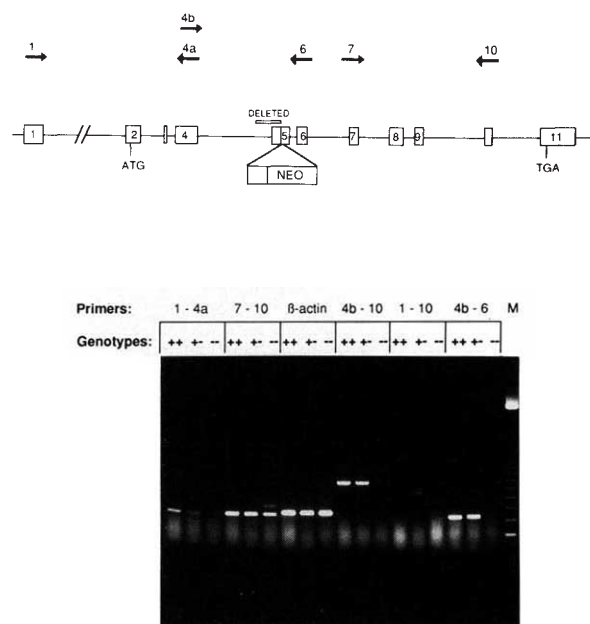


FIG. 4 Homozygous mice do not express intact p53 RNA. RNA-PCR was performed on spleen RNA from wild-type, heterozygote and homozygote mice. The PCR primers used and their relative locations within the p53 gene are indicated by labelled arrows in the top diagram. Positions of the *neo* cassette insertion and exon 5 deletion within the targeted p53 gene are illustrated. The bottom panel shows the RNA-PCR products subjected to agarose gel electrophoresis and visualized by ultraviolet illumination in the presence of ethidium bromide. PCR primer pairs and the mouse genotype are indicated at the top of the gel for each amplified RNA sample. Lane 'M' shows a 123-bp ladder as size calibration marker (BRL). β -Actin primers were used as controls to show that the amount of RNA in each reaction did not vary greatly and that the homozygote RNAs did not contain an inhibitor of the amplification process.

METHODS. Kidney, spleen, and testes of mice were homogenized and total RNA was then extracted from the homogenates using the RNeasy B kit (Qiagen) according to the manufacturer's specifications. Total RNA (2-4 μ g) from each tissue were used for cDNA synthesis followed by PCR amplification using the Perkin-Elmer RNA-PCR kit. Reaction conditions were as specified by the manufacturer. An oligo(dT) primer was used to initiate cDNA synthesis. Thirty PCR cycles (30 s at 94 $^{\circ}$ C; 1 min at 60 $^{\circ}$ C; 2 min at 72 $^{\circ}$ C) were run and the amplification products visualized after electrophoresis on agarose gels (1.2%) under ultraviolet illumination in the presence of ethidium bromide. Gels were blotted and hybridized to a labelled p53 probe to verify that the bands were p53-specific. The sequences of the mouse p53 primers were: primer 1 (exon 1; bp 11) 5'-CACGTGCTCACCCTGGCTAAAAGT-3'; primer 4a (exon 4; bp 300) 5'-TGGGGCAGCAACAGATCGTCAT-3'; primer 4b (exon 4, bp 497) 5'-GGGACAGCCAAGTCTGTTATGTGC-3'; primer 6 (exon 6, bp 774) 5'-CTGTCTTCCAGATACTCGGGATAC-3'; primer 7 (exon 7, bp 881) 5'-GGCAT-GAAGTGGAGGCCTATCCTTACCAT-3'; primer 10 (exon 10, bp 1177) 5'-CTCCC-GGAACATCTCGAAGC-3'.

TABLE 1 Tumours in p53-deficient mice

Case	Sex	Genotype	Age (weeks)	Histological type	Anatomic site
1	M	homozygote	18	undifferentiated sarcoma	subcutis
2	M	homozygote	8	gonadoblastoma	testis
3	M	homozygote	18	malignant lymphoma	thymus
4	F	homozygote	9	haemangiosarcoma (2)	subcutis/liver
5	F	homozygote	15	malignant lymphoma	generalized
6	M	homozygote	24	malignant lymphoma	generalized
7	M	homozygote	24	undifferentiated sarcoma	vertebra
8	M	homozygote	15	haemangiosarcoma	subcutis
9	M	homozygote	24	haemangiosarcoma	subcutis
10	M	homozygote	16	malignant lymphoma	generalized
11	F	homozygote	15	osteosarcoma	pelvis
12	F	homozygote	18	malignant lymphoma	generalized
13	M	homozygote	14	haemangioma	heart
14	F	homozygote	14	malignant lymphoma	thymus
15	F	homozygote	15	malignant lymphoma	generalized
16	M	homozygote	16	mammary adenocarcinoma	mammary gland
17	M	homozygote	17	malignant lymphoma	generalized
18	M	homozygote	17	haemangiosarcoma	subcutis
19	M	homozygote	21	haemangioma	muscle
20	M	homozygote	22	malignant lymphoma	thymus
21	M	homozygote	22	malignant lymphoma	thymus
22	M	homozygote	25	malignant lymphoma	generalized
23	M	homozygote	19	alveolar/bronchiolar adenoma	lung
24	M	homozygote	21	undifferentiated sarcoma	lumbar muscle
25	F	homozygote	26	malignant lymphoma	generalized
26	M	homozygote	24	malignant lymphoma	generalized
27	M	heterozygote	24	malignant lymphoma	generalized
28	M	heterozygote	37	haemangiosarcoma (2)	generalized
29	M	chimaeric	60	haemangioma	generalized
30	M	chimaeric	60	embryonal carcinoma	generalized
				malignant Schwannoma	salivary gland
				haemangiosarcoma	subcutis
				choriocarcinoma	testis
				Leydig cell tumour	testis

* One phenotypically male chimaeric mouse had a choriocarcinoma surrounded by recognizable ovarian tissue.

homozygote mice (Fig. 4, top). The results of these RNA-PCR assays indicated that an intact p53 mRNA was synthesized only in the wild-type and heterozygote mice. Representative amplification products are shown for total spleen RNA (Fig. 4, bottom). Primer pairs specific for transcripts initiated at the p53 promoter that cross the *neo* insertion site (primer pairs 4b-6, 4b-10, and 1-10) were unable to amplify a PCR product from samples derived from the homozygote animals, indicating their failure to produce a message initiated at the p53 promoter that extended beyond exon 4. Exon 7-10 primers generated a strong PCR product from the same homozygote cDNAs. This product was probably derived from a hybrid *neo*-p53 message initiated by the inserted *Pol*III promoter; such a transcript was expected because our targeting construct was designed to generate a *neo* fusion RNA which used the p53 polyadenylation signal. Exon 1-4 primers generated a PCR band from the homozygote cDNAs that was very faint compared with that from wild-type cDNAs. We suggest that transcription from the endogenous p53 promoter in homozygotes could result in an unstable truncated transcript because of the *neo* cassette insertion.

As a final characterization of the p53 null allele, we used metabolic labelling and protein immunoblotting to assay p53 protein expression in secondary fibroblasts derived from normal, heterozygote and homozygote mice. For protein immunoblotting extracts of fibroblast cultures were immunoprecipitated using a

pool of monoclonal antibodies specific for either the amino terminus of the p53 protein (Fig. 5a, lanes 2, 5, 8) or for the entire p53 molecule (Fig. 5a, lanes 3, 6, 9). Intact p53 was detected in fibroblasts from normal and heterozygote mice (Fig. 5a, lanes 2, 3, 5, 6) but was not detectable in the homozygote fibroblasts (Fig. 5a, lanes 8, 9). Metabolic labelling experiments confirmed the results from protein immunoblotting (Fig. 5b). There is no evidence of a truncated amino-specific p53 protein that might have been generated by insertion of the *neo* gene into exon 5 of p53 in the targeting construct and whose predicted relative molecular mass would be 17,000-20,000 (Fig. 5b, lanes 5, 6, 8, 9). The low level of p53 detected in these fibroblasts is consistent with that expected for wild-type p53 and contrasts with the raised levels of p53 commonly observed for mutant forms of the protein. Together with our results from Southern blotting and RNA-PCR, these data indicate that neither a truncated nor a full-length p53 protein is produced in cells from homozygous mice.

Tumour susceptibility

The mice carrying mutant p53 alleles enabled us to test the effects of the loss of a tumour-suppressor gene on tumour development. The original chimaeras, heterozygote and homozygote p53-deficient mice, and normal littermates (with wild-type p53 genes) were monitored for spontaneous neo-

plasms. Wild-type mice (95 animals) by the age of 9 months failed to develop any tumour or illness. One chimaeric mouse (aged 14 months) developed two different primary tumours, an osteogenic sarcoma and an ovarian choriocarcinoma. A second chimaeric mouse (also 14 months old) had a Leydig-cell tumour of the testis. By 9 months only 2 of 96 heterozygote animals have developed tumours. One heterozygote mouse developed an embryonal carcinoma of the testis at 5 months and another developed a malignant lymphoma at the age of 9 months.

A dramatic susceptibility to the development of multiple types of cancer was observed in the homozygotes. Out of a study population of 35 homozygote animals, 26 (74%) developed obvious neoplasms by 6 months of age. Some tumours appeared very early (before 10 weeks of age), and tumour occurrence increased rapidly between 15 and 25 weeks of age. For those homozygous mice developing tumours, the mean time to tumour appearance was 20 weeks. Of the 35 homozygotes, six mice died from unknown causes. When we examined these animals we found no overt cancers. Four non-tumour deaths seemed to be related to unresolved infections, as the mice had abscesses, gastroenteritis or myocarditis. Two mice died without obvious tumours, but tissue autolysis prevented a thorough pathological examination.

The homozygote tumours were analysed in detail by histopathology (Table 1). These tumours were remarkable with respect to the variety of cell types involved, their relatively early onset, and their apparent malignancy. Multiple primary neoplasms of different cell types of origin were observed in 9 of the 26 tumour-bearing homozygote mice (Table 1). The tumours most frequently observed were malignant lymphomas affecting 20 of 26 animals, but sarcomas also occurred with some frequency. Most lymphomas involved the thymus and other major visceral organs (heart, lung, spleen, liver, kidney, brain)

and were of the lymphoblastic cell type (Fig. 6). Of the sarcomas, seven were haemangiosarcomas, three were undifferentiated sarcomas, and one was an osteosarcoma. One of the undifferentiated sarcomas arose from the subcutis, invading the abdominal wall into the body cavity, and another replaced the second lumbar vertebra, causing hind-limb paralysis. One female mouse developed a mammary adenocarcinoma. Benign neoplasms such as haemangiomas were also observed.

Discussion

We used a homologous recombination strategy to introduce a null mutation into the p53 gene of embryonic stem cells. The strategy used a vector with a very strong selection for rare gene replacement events. Although the targeting efficiency of two correct homologous recombination events in 2×10^8 electroporated embryonic stem cells was relatively low, the removal of the polyadenylation signal from the *neo* cassette in the targeting vector resulted in only 100 total colonies that had to be screened after G418 and FIAU selection. The disrupted p53 allele was successfully transmitted in the germ line of mice. Crosses of mice carrying one null allele resulted in a homozygous (null) genotype in ~25% of the offspring. The mice appeared healthy and lacked any observable gross defects.

The apparently normal development of mice nullizygous for p53 indicates that p53 is dispensable for mouse development. One explanation for the normal development of the homozygous mice is that a truncated p53 protein may provide necessary functions. However, our targeting strategy produced a profound disruption and deletion of the p53 gene that was unlikely to generate a functional protein. Southern and northern hybridization, RNA-PCR, protein immunoprecipitation and immunoblot experiments verified that an intact p53 was not being synthesized in the homozygote animals. The immunoprecipitation

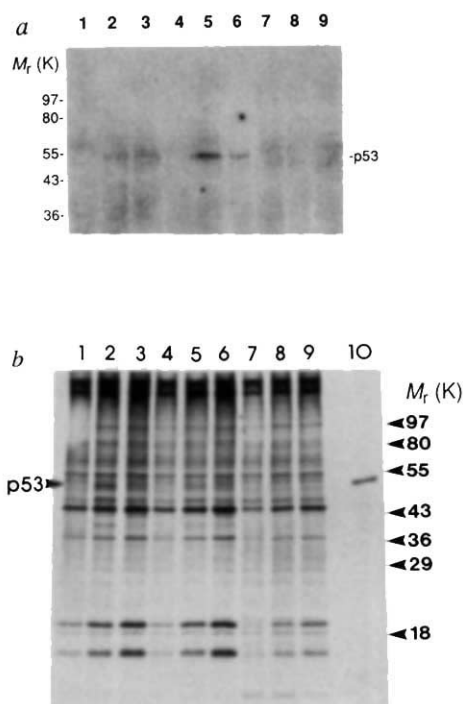


FIG. 5 Fibroblasts from homozygous mice do not express detectable quantities of intact p53 protein. *a*, Detection by protein immunoblot. Secondary embryo fibroblast cultures derived from wild-type (lanes 4–6), heterozygote (lanes 1–3), and homozygote embryos (lanes 7–9) were extracted with detergents, and the clarified extracts characterized by immunoprecipitation, SDS-polyacrylamide gel electrophoresis and immunoblot analysis. Monoclonal antibodies used for immunoprecipitation of extracted proteins included control antibody m73 (specific for adenovirus E1A) (lanes 1, 4, and 7), a mixture of p53 monoclonal antibodies PAb242, PAb246 and PAb248 (lanes 2, 5, and 8), or a mixture of PAb421, RA3.2C2 and 200.37 (lanes 3, 6, and 9). The migration of pre-stained molecular weight (in thousands) markers is shown at the left. Note the absence of intact or truncated p53 protein in extracts derived from homozygote fibroblasts. *b*, Detection by metabolic labelling. Secondary embryo fibroblast cultures were starved for 30 min in methionine-free medium and were then labelled for 30 min in medium containing $200 \mu\text{Ci ml}^{-1}$ [^{35}S]methionine. Cells proteins were then extracted and characterized by immunoprecipitation, SDS-PAGE and autoradiography. Monoclonal antibodies used for immunoprecipitation included m73 (lanes 1, 4, 7), a pool of monoclonal antibodies specific for the amino portion of p53 (PAb242, PAb246, PAb248, 200.47; lanes 2, 5, 8), and PAb248 (lanes 3, 6, 9). The migration of *in vitro*-translated, ^{35}S -labelled p53 is shown in lane 10. Molecular weight (in thousands) markers are shown at the right. Note the absence of p53 in cells derived from homozygote mice (lanes 7–9). Bands migrating at 48 K and 40 K (lanes 2, 3, 5, 6) either represent degradation products of p53 or are co-precipitated with the p53 protein. The faint band observed at 16K in lane 8 is considered nonspecific as it is also present in the control lanes after a longer exposure of the film. **METHODS.** Equivalent numbers of secondary embryo fibroblasts in log-phase growth were rinsed with phosphate-buffered saline and their proteins extracted⁴⁵. Clarified extracts were then immunoprecipitated with antibodies, and the immunoprecipitated proteins separated by electrophoresis on 12% SDS-polyacrylamide gels⁴⁶. Proteins were transferred to nylon membranes and immunoblotted with primary antibody and ^{125}I -labelled protein A as described⁴⁷. Monoclonal antibodies used for immunoprecipitation were m73 (ref. 48), PAb421 (ref. 49), Pab242 (ref. 50), PAb246 (ref. 50), PAb248 (ref. 50), 200.47 (ref. 51), and RA3.2C2 (ref. 52). All of these monoclonal antibodies except for PAb421 and m73 react with amino-terminal portions of p53. A mixture of all p53 monoclonal antibodies (each diluted 1:200) was used as the blotting antibody.

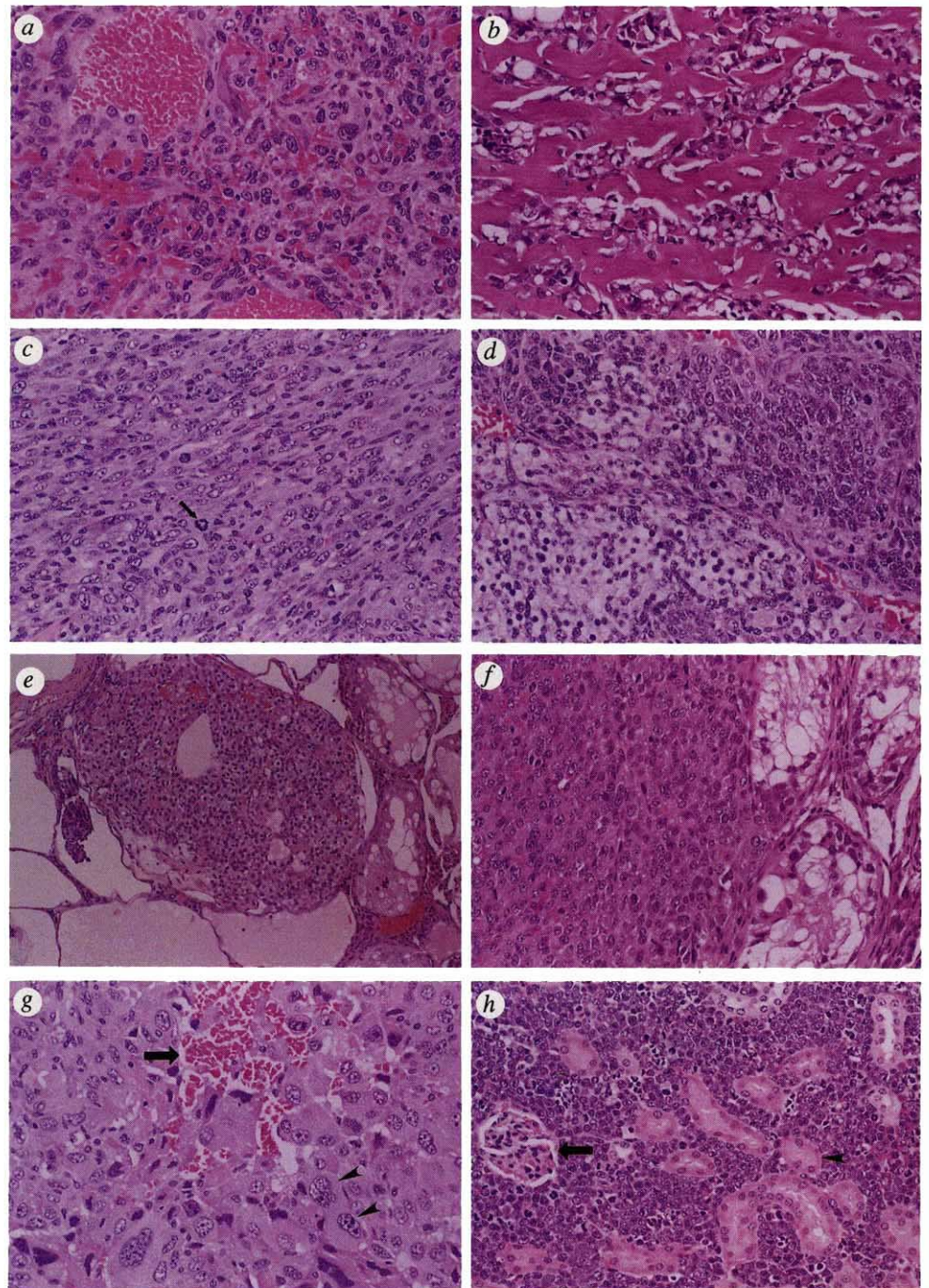
experiments indicated that no sizeable truncated p53 product was detectable in the homozygotes. Given the apparent function of p53 in cell-cycle regulation²¹⁻²⁸, a nullizygous p53 state would be expected to deregulate the cell cycle in many cell types during differentiation and development, resulting in aberrant morphogenesis leading to an embryonic lethal condition. It is possible that during development cells may use other proteins or pathways to accomplish the functions normally provided by p53. Alternatively, perhaps p53 does not have a role in cell cycle control during development. Although mouse tumour cells lacking p53 have been observed³⁷, we had not anticipated that the entire organism could survive without the protein.

Although the p53-negative homozygous mice appeared developmentally normal, they were clearly susceptible to spontaneous tumour formation (naturally occurring tumours are rare in mice of less than 6 months of age^{38,39}). The development of

multiple types of tumours by 6 months of age in nearly 75% of the p53 homozygote mice provides *in vivo* confirmation of the tumour-suppressor property of normal p53 previously established *in vitro*^{9,10}. The tumour incidence in the homozygous mice is considerably higher than the 20% incidence observed by 18 months of age in mice carrying a mutant p53 transgene⁴⁰. In the transgenic mouse, the presence of two wild-type p53 alleles (each with tumour suppressor activity) may be an important modifying factor. (Additionally, tissue-specific expression levels of the mutant transgene may dictate the rates and types of tumour development.) The variety of tumours identified in the homozygote mice demonstrate that p53 is involved in tumorigenesis in many tissues and cell types. But although we found a number of different tumour types, the most frequently occurring neoplasm was malignant lymphoma. More than 70% of the homozygous mice developed this cancer. This predom-

FIG. 6 Histopathology of tumours from p53-deficient mice. *a*, Subcutaneous haemangiosarcoma. Note attempts to form new blood vessels and larger blood-filled spaces. *b*, Well-differentiated sacral osteosarcoma with numerous pink islands of osteoid. *c*, Subcutaneous undifferentiated sarcoma with pleomorphic spindle cells, high mitotic index, and atypical mitoses (arrow). *d*, Testicular gonadoblastoma with germ-cell component (seminoma) in upper right and stromal component (Leydig cells) in lower left. *e*, Small, well-differentiated Leydig cell tumour of the testis. Note atrophic seminiferous tubules to the right of the tumour. *f*, Solid embryonal carcinoma of the testis. Atrophic seminiferous tubules can be seen to the right of the tumour. *g*, Choriocarcinoma with trophoblastic giant cells⁵³ (arrowheads) and intercellular haemorrhage (large arrow). *h*, Lymphoblastic malignant lymphoma in the kidney. Note solid sheets of lymphoma cells infiltrating between glomeruli (large arrow) and renal tubules (arrowheads) in the cortex.

METHODS. Selected tissues were fixed in 10% neutral buffered formalin and subsequently dehydrated through ascending grades of alcohol, cleared in xylene, and infiltrated with paraffin wax using a Miles VIP-3000 tissue processor. The paraffin blocks were sectioned on a Leitz 1512 rotary microtome at 4 μ m, and sections were retrieved from the water bath. Sections were stained with a haematoxylin and eosin, dehydrated, and cleared using a Fisher Code-On automated system. Sections were viewed with an Olympus BHS-2 microscope.



ance of malignant lymphomas in our mice contrasts with the range of tumours observed in the p53 transgenic mice, in which lymphomas occurred in only about a third of the tumour-bearing mice. The predisposition to particular types of tumours in mice may be partly strain-dependent. The p53 transgenic mice were outbred CD-1 mice. Our p53-deficient mice had genetic components from the C57BL/6 strain (~75%) and from the 129/Sv strain (~25%). Interestingly, C57BL/6 mice develop lymphomas at a high rate, but at a mean age of 27 months⁴¹. Lymphomagenesis in our homozygote mice occurred with a mean age to lymphoma of about 5 months. In addition, 129/Sv mice have a relatively high incidence of testicular tumours⁴², and we noted this type of tumour in several of the p53-deficient mice. Therefore, loss of p53 may greatly accelerate a prior tumour predisposition. We propose that the range of tumour types in p53-deficient mice will vary according to the genetic background. The increased tumorigenesis in these mice clearly demonstrates that the loss of p53 function is sufficient to genetically predispose animals to spontaneous tumour development; a p53 oncogenic point mutation is not required as an initiating event. The sporadic development of spontaneous tumours indicates that other genetic or epigenetic events in addition to the loss of p53 are required for the development of fully malignant neoplasms.

Tumour development in these homozygous mice is reminiscent of that in Li-Fraumeni families^{20,21} in that a p53-related pathway of transformation leads to the development of a variety of tumour types. But Li-Fraumeni patients inherit one wild-type p53 gene and one mutant gene, a situation more analogous to the heterozygote mice bearing a single null p53 allele (which developed spontaneous tumours at a modest rate). The majority of Li-Fraumeni families studied so far have p53 alleles with a single point mutation in exon 7. Therefore it remains uncertain whether our heterozygotes with a single null allele develop tumours through a molecular pathway(s) similar to Li-Fraumeni patients with a point mutant allele. But one Li-Fraumeni family has been found to have a p53 allele with a frameshift mutation in exon 5 (ref. 43). Such an allele would generate a truncated

(and presumably inactive) p53 protein roughly similar to that hypothesized for our p53-deficient mice. In this case the p53-deficient mice may represent a genetically accurate animal model for a human inherited cancer syndrome.

A significant number of p53-negative homozygous mice died without obvious tumours (tissue autolysis prevented a complete pathological analysis in some cases). Several of the deaths were associated with unresolved infections, suggesting the presence of a defect of the immune system in the homozygotes. To address this issue, preliminary examinations of thymus and spleen lymphoid subsets by flow cytometry indicate that homozygotes are not deficient in the number of total B cells or CD4⁺ or CD8⁺ T cells (S. Rich, unpublished data). Antibody isotyping experiments demonstrated normal levels of IgA, IgM and IgG subclasses in wild-type, homozygote, and heterozygote mice (data not shown). Finally, homozygotes were able to mount an apparently normal humoral immune response to a test antigen (data not shown). These initial studies suggest that if the homozygotes have an immune defect, it may be a subtle one, and further functional analyses will be necessary to uncover it. If immune defects are present in the homozygotes, they could contribute to the accelerated rate of tumour development. But we believe it is unlikely that immune defects could be the sole cause of the homozygote tumours, based on the behaviour of the homozygote cells in culture. Embryo fibroblasts derived from homozygote animals consistently immortalize when passaged at moderate density, whereas heterozygote and wild-type embryo fibroblasts passaged under the same conditions consistently stop dividing and die at 8–15 passages (M.H., unpublished data). These differences in *in vitro* growth phenotype argue that the homozygote cells in *in vivo* may be intrinsically more predisposed to neoplastic transformation, regardless of external parameters such as immune system defectiveness.

We believe that the p53-deficient mouse will be a useful model for studying the role of the p53 tumour-suppressor gene in tumorigenesis. The increased susceptibility of these mice to a wide range of tumours makes them valuable for testing suspected carcinogens and cancer therapeutic agents. □

Received 15 November 1991; accepted 20 February 1992.

- Lane, D. P. & Crawford, L. V. *Nature* **278**, 261–263 (1979).
- Linzer, D. I. H. & Levine, A. J. *Cell* **17**, 43–52 (1979).
- Sarnow, P., Ho, Y. S., Williams, J. & Levine, A. J. *Cell* **28**, 387–394 (1982).
- Werness, B. A., Levine, A. J. & Howley, P. M. *Science* **248**, 76–79 (1990).
- Eliyah, D. et al. *Nature* **312**, 646–649 (1984).
- Parada, L. F. et al. *Nature* **312**, 649–651 (1984).
- Jenkins, J. R., Rudge, K. & Currie, G. A. *Nature* **312**, 651–654 (1984).
- Jenkins, J. R., Rudge, K., Chumakov, P. & Currie, G. A. *Nature* **317**, 816–818 (1985).
- Finlay, C. A., Hinds, P. W. & Levine, A. J. *Cell* **57**, 1083–1093 (1989).
- Eliyah, D. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8763–8767 (1989).
- Mercer, W. F. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6166–6170 (1990).
- Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. V. & Vogelstein, B. *Science* **249**, 912–915 (1990).
- Diller, L. et al. *Molec. cell. Biol.* **10**, 5772–5781 (1990).
- Michalovitz, D., Halevy, O. & Oren, M. *Cell* **62**, 671–680 (1990).
- Chen, P.-L., Chen, Y., Brookstein, R. & Lee, W.-H. *Science* **250**, 1576–1580 (1990).
- Mowat, M., Cheng, A., Kicunma, N., Bernstein, A. & Benchimol, S. *Nature* **61**, 2777–2781 (1985).
- Ben David, Y., Pradeaux, V. R., Chow, V., Benchimol, S. & Bernstein, A. *Oncogene* **3**, 179–185 (1988).
- Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. *Science* **253**, 49–53 (1991).
- Vogelstein, B. *Nature* **348**, 681–682 (1990).
- Malkin, D. et al. *Science* **250**, 1233–1238 (1990).
- Srivastava, S., Zou, Z., Pirolo, K., Blattner, W. & Chang, E. H. *Nature* **348**, 757–749 (1990).
- Stürzbecher, H.-W. et al. *Oncogene* **5**, 795–801 (1990).
- Bischopf, J. R., Friedman, P. N., Marshak, D. R., Privet, C. & Beach, D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4766–4770 (1990).
- Shaulsky, G., Ben-Ze'ev, A. & Rotter, V. *Oncogene* **5**, 1707–1711 (1990).
- Marshall, C. J. *Cell* **64**, 313–326 (1990).
- Levine, A. J., Momand, J. & Finlay, C. A. *Nature* **351**, 453–456 (1991).
- Gannon, J. V. & Lane, D. P. *Nature* **329**, 456–458 (1987).
- Braithwaite, A. W. et al. *Nature* **329**, 458–460 (1987).
- Fields, S. & Jang, S. K. *Science* **249**, 1046–1049 (1990).
- Raycroft, L., Wu, H. & Lozano, G. *Science* **249**, 1049–1051 (1990).
- Rogel, A., Popliker, M., Webb, C. & Oren, M. *Molec. cell. Biol.* **5**, 2851–2855 (1985).
- Mansour, S. L., Thomas, K. R. & Capecci, M. R. *Nature* **336**, 348–352 (1988).
- Soriano, P., Montgomery, C., Geske, R. & Bradley, A. *Cell* **64**, 693–702 (1991).
- Soussi, T., de Fromental, C. C. & May, P. *Oncogene* **5**, 945–952 (1990).
- McMahon, A. & Bradley, A. *Cell* **62**, 1073–1085 (1990).
- Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 113–151 (IRL, Oxford, 1987).
- Wolf, D. & Rotter, V. *Molec. cell. Biol.* **4**, 1402–1410 (1984).
- Maita, K. et al. *Toxicol. Pathol.* **16**, 340–349 (1988).
- Squire, R. A. et al. in *Pathology of Laboratory Animals Vol. II* (eds Benirschke, K. et al.) 1054–1055 (Springer, New York, 1978).
- Lavigne, A. et al. *Molec. cell. Biol.* **9**, 3982–3991 (1989).
- Stutman, O. *Fed. Proc. Fed. Am. Soc. exp. Biol.* **33**, 2028–2032 (1974).
- Stevens, L. C. *J. natn. Cancer Inst.* **50**, 235–242 (1974).
- Yandell, D. W. *Proc. Am. Assoc. Cancer Res.* **32**, 459–460 (1991).
- Bienz, B., Zakut-Houri, R., Libresco, S., Givol, D. & Oren, M. *EMBO J.* **3**, 2179–2183 (1984).
- Jarvis, D. L., Lanford, R. E. & Butel, J. S. *Virology* **134**, 168–176 (1984).
- Sepulveda, A. R. et al. *Cancer Res.* **49**, 6108–6117 (1989).
- Slagle, B. L., Lanford, R. E., Medina, D. & Butel, J. S. *Cancer Res.* **42**, 2155–2162 (1984).
- Harlow, E., Franza, B. R. Jr & Schley, C. J. *Virology* **55**, 533–546 (1985).
- Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. *J. Virol.* **39**, 861–869 (1981).
- Yewdell, J. W., Gannon, J. V. & Lane, D. P. *J. Virol.* **59**, 444–452 (1986).
- Dippold, W. G., Jay, G., DeLeo, A. B., Khoury, G. & Old, L. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1695–1699 (1981).
- Coffman, R. L. & Weissman, I. K. *J. exp. Med.* **153**, 269–279 (1981).
- Alison, R. H., Lewis, D. J. & Montgomery, C. A. *Vet. Path.* **24**, 226–230 (1987).

ACKNOWLEDGEMENTS. We thank W. Fang, S. Nyaga, M. Ozbun, S. Jackson, J. D. Wallace, A. Davis, S. Rich, M. Wims and R. Geske for advice and technical assistance. We thank P. Soriano, M. Oren and G. G. Lozano for plasmids. This research was supported by the Texas Advanced Technology Program and by the American Cancer Society.

X-ray detection of Nova Herculis 1991 five days after optical outburst

H. M. Lloyd*, T. J. O'Brien*, M. F. Bode*†, P. Predehl‡, J. H. M. M. Schmitt‡, J. Trümper‡, M. G. Watson§ & K. A. Pounds§

* Department of Physics and Astronomy, Lancashire Polytechnic, Preston PR1 2TQ, UK

‡ Max-Planck-Institut für Extraterrestrische Physik, Karl-Schwarzschild-Strasse 1, D-8046 Garching, Germany

§ Department of Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK

CLASSICAL nova outbursts are thought to occur in binary systems in which a white dwarf accretes material from a main-sequence dwarf. The outburst is due to thermonuclear runaway in the accreted material, and results in the ejection of about 10^{-5} – 10^{-4} solar masses of material at velocities of up to several thousand kilometres per second (ref. 1). Previous X-ray observations of classical novae in the early stages of outburst have resulted only in upper limits to the X-ray flux^{2–4}. Here we report a positive detection by the Rosat satellite of X-ray emission from Nova Herculis 1991 five days after its optical discovery. Standard nova models predict X-ray emission to arise directly from nuclear burning on the surface of the white dwarf, and suggest that X-rays should not be seen until later in the outburst⁵. We argue that the emission from Nova Her 1991 comes from hot, shocked circumstellar material, which may be the ejected material itself or pre-existing circumstellar matter. In either case, however, the required density of material is higher than models of nova binary systems would suggest.

Nova Herculis 1991 was discovered independently by United Kingdom and Japanese amateur astronomers⁶ on 1991 March 24–25. Estimated visual magnitudes of $V \approx 5$ at discovery suggest that Nova Her 1991 is one of the brightest novae at optical maximum of recent decades. It is unclear whether the nova was caught at optical maximum or just after this. It is clear, however, that Nova Her 1991 is the fastest nova known, taking only about a day to decline two magnitudes from its observed peak brightness (see Fig. 1). In the weeks following discovery, Nova Her 1991 was extensively observed at several wavelengths. In the radio regime, the nova was detected with the Very Large Array on 1991 April 5.43 UT at millijansky levels at gigahertz frequencies⁷. Optical observations indicate broad emission lines; the $H\alpha$ emission on March 29.5 had a full width at half maximum of $\sim 4,500$ km s⁻¹, with a spectrum similar to that of V1500 Cyg in outburst, suggesting a possible magnetic white-dwarf primary⁸. Infrared observations showed an increase in K and L brightness⁹ up to April 1 that was consistent with the formation of dust very early in the outburst or the irradiation of a previously existing dust shell. Additional evidence for the presence of dust has been provided by optical spectra obtained on April 4 and 5 showing the broad $H\alpha$ emission line split into five components, with the relative heights of the blue and red peaks suggesting absorption by a forming dust shell¹⁰. Further infrared observations¹¹ showed a steady decline in J, H and K magnitudes during April 5–22.

Here we present X-ray observations of Nova Her 1991 obtained on March 30.46 with the Position Sensitive Proportional Counter (PSPC) and X-ray Telescope (XRT) aboard Rosat¹², five days after discovery¹³. Only three classical novae have been observed in X-rays near optical maximum: V1500 Cyg with SAS-3 (ref. 2), NQ Vul with Ariel V (ref. 3) and PW

Vul with Exosat⁴. In each case only an upper limit to the X-ray flux could be derived. X-ray emission has been observed and detected later in the decline from optical maximum, however; for example GQ Mus (1983), PW Vul (1984) and QU Vul (1984) showed soft X-ray emission consistent in all three novae with nuclear burning on the surface of the white dwarf⁴. All other X-ray observations of classical novae were made several years after outburst⁵. The Rosat PSPC observation of Nova Her 1991 on March 30.46 UT resulted in a firm detection with a count rate of 0.16 ± 0.01 s⁻¹ in a 1,235-s exposure. Examination of the PSPC survey exposure obtained between 1990 September 25 and 28 showed no source at the same position with an upper limit to the count rate of 0.01 s⁻¹.

Figure 2 shows the PSPC spectral data along with a best-fit Raymond-Smith¹⁴ thermal spectrum (see later). Note that the spectrum has a low-energy cut-off at a photon energy of ~ 1 keV, suggesting that the source is highly absorbed. Although only 194 photons were observed during the integration time, the gross spectral properties of the source strongly suggest that the emission is from a hot plasma. A large column rules out black-body emission from the optically thick wind of the nova as the source of the observed X-rays (the effective temperature of the wind at 4 magnitudes below optical maximum would be $\sim 5 \times 10^4$ K, ref. 15). Similarly, we can also rule out hotter black-body emission ($T \approx 5 \times 10^5$ K) from the surface of the white dwarf, where thermonuclear burning can continue long after outburst¹⁶, a model which has been successfully applied to observations of earlier novae. As Fig. 3 shows, the spectral properties are, however, entirely consistent with the flat spectrum of high-temperature thermal emission observed through a large absorbing column (the high-energy cut-off is due to the instrumental response). The possibility of using Rosat to distinguish between these possible X-ray emission mechanisms in nova outbursts was first pointed out by Ögelman¹⁷.

We have therefore fitted the spectral data with a Raymond-Smith thermal spectrum under the assumption of solar abundances of the heavy elements. We note that using abundances appropriate to the most enriched nova ejecta would not alter our density estimates (see later) by more than a factor of three. A fit exists at $kT \approx 10$ keV with $N_H \approx 3.5 \times 10^{21}$ cm⁻². Given the limited spectral bandpass of the PSPC, it is impossible to distinguish between fits at these high temperatures, so it seems reasonable to hold the temperature fixed at $kT = 10$ keV. This yields a best-fit column density of $N_H = (3.4 \pm 1.6) \times 10^{21}$ cm⁻² at a reduced χ^2 of 2.13. There also exists a low-temperature fit with

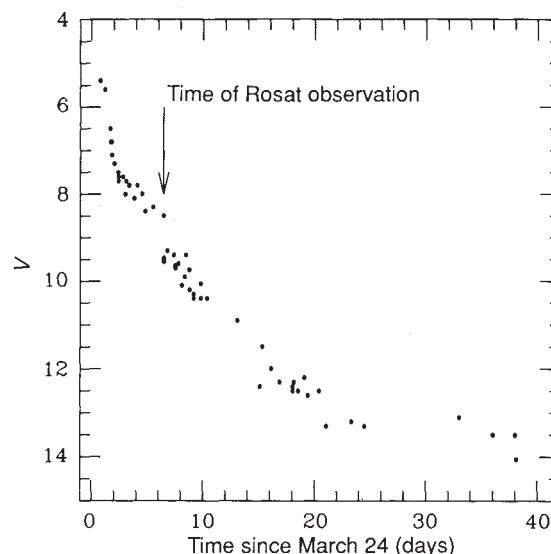


FIG. 1 Visual light curve of Nova Her 1991 compiled from refs 6, 21–30.

† Present address: Physical and Chemical Sciences, Liverpool Polytechnic, Byrom Street, Liverpool L3 3AF, UK.

$kT = 0.75 \pm 0.25$ keV and $N_H = (1.6 \pm 0.2) \times 10^{22} \text{ cm}^{-2}$ at $\chi^2 = 2.2$. The emission measure $E_{0.75}$ of the material derived from the spectral fit at $kT = 0.75$ keV is given by $E_{0.75} = 2.3 \times 10^{56} d^2 \text{ cm}^{-5}$, where d is the distance to the object in kiloparsecs. For the high-temperature fit, we find that the emission measure E_{10} is given by $E_{10} = 7.6 \times 10^{55} d^2 \text{ cm}^{-5}$. The fluxes in the 0.2–2.4 keV band for the 0.75-keV and 10-keV fits are $7 \times 10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1}$ and $5.1 \times 10^{-12} \text{ erg cm}^{-2} \text{ s}^{-1}$ respectively. Note that the emission observed from Nova Her 1991 five days after discovery is consistent with the upper limit on the Exosat count rate from PW Vul eight days after optical maximum⁴, assuming that the emission in both cases is from a hot plasma.

Determining the distance to the object is, of course, difficult; the most reliable methods, using expansion parallax, are not possible at this early stage. We can, however, estimate the distance to Nova Her 1991 to be $\sim 10 \pm 4$ kpc, on the basis of four independent methods: the empirical law relating the absolute magnitude of a nova at maximum to its speed class¹⁸; the empirical fact that all novae have roughly the same absolute magnitude 15 days after optical maximum¹⁹; the derived black-body angular diameter of the optically thick photosphere at optical maximum (assumed to occur 1 day from outburst and have an effective temperature of 10^4 K); and the black-body angular diameter of the dust shell 17 days after optical maximum with a dust temperature $T_{\text{dust}} = 1,200$ K (ref. 19). For the optical methods, we have assumed a visual extinction $A_V = 1.5$ (an upper limit of $E_{B-V} = 0.5$ has been derived from preliminary reduction of spectra from the International Ultraviolet Explorer; S. Starrfield, personal communication) and $V_{\text{max}} = 5.4$. We note that this is consistent with the best value of N_H for the higher-temperature spectral fit²⁰. For the black-body angular diameter methods, an expansion velocity of $4,000 \text{ km s}^{-1}$ has also been assumed. This distance implies an absolute magnitude at maximum of $M_V = -11.1$, comparable to the absolute magnitude at maximum of V1500 Cyg ($M_V = -10.7$), which was previously the fastest known nova.

What then is the origin of the observed X-ray emission of Nova Her 1991? Having eliminated remnant burning or pseudo-photospheric emission as discussed above, a more plausible explanation is that the emitting material is shock-heated to X-ray temperatures by interaction of the high-velocity ejecta with some pre-existing material surrounding the nova. For a strong shock moving at a velocity comparable to that of the ejecta ($\sim 4,000 \text{ km s}^{-1}$) into the ambient medium, the post-shock temperature is given by $kT = 18.75$ keV. The hot, shocked, circumstellar material in turn drives a reverse shock into the colder ejecta. The ratio of the post-shock temperatures at the two shocks is equal to the inverse of the ratio of the pre-shock densities, so a certain minimum density is required in order for the ejecta

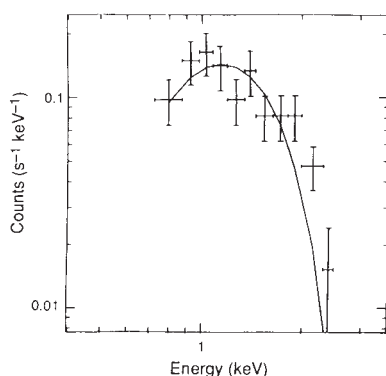


FIG. 2 PSPC spectral data for the observation of Nova Her taken at March 30.46 UT. The solid line is the best-fit Raymond-Smith thermal spectrum with $kT = 10$ keV and a column density $N_H = (3.4 \pm 1.6) \times 10^{21} \text{ cm}^{-2}$.

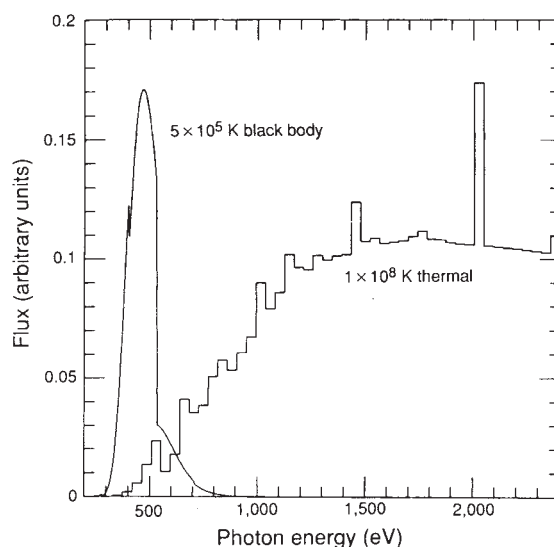


FIG. 3 Comparison of the spectra of the possible sources of the X-ray emission observed in Nova Her 1991, with arbitrary normalization. The spectra shown are black-body emission at 5×10^5 K expected from thermonuclear burning on the white dwarf and the thermal emission from a plasma with solar abundances at a temperature of 10^8 K. Both spectra are subject to an absorbing column of $N_H = 3.5 \times 10^{21} \text{ cm}^{-2}$. The cooler black-body emission ($T \approx 5 \times 10^4$ K) from the optically thick wind of the nova is severely absorbed by the interstellar column and is not shown here. The spectrum of the source can be due only to the high-temperature thermal emission.

to be shocked to X-ray temperatures. There are therefore two possible sources of emission, the shocked ambient material and the shocked ejecta. We estimate from the emission measures that a circumstellar density of $\sim 10^{-18}$ to $10^{-17} \text{ g cm}^{-3}$ is required to provide emission either from the reverse-shocked ejecta at $kT = 0.75$ keV or from the forward-shocked ambient material at the higher temperature, although in each case the other source of emission contributes at a non-negligible level. We note that these derived densities depend inversely on the assumed distance and will not therefore be more than an order of magnitude in error. The implication is, however, that whether the emission arises from the forward shock or the reverse-shocked ejecta, the outburst is occurring in a relatively dense medium.

Our observation of X-rays after five days is unexpected, as standard models of classical novae suggest that X-rays would not be seen until much later in the outburst (see ref. 5), the emission being due to nuclear burning on the surface of the white dwarf. Our suggested explanation is that the ejecta interact with previously existing material surrounding the nova. But little circumstellar material is expected in a binary system consisting of a main-sequence dwarf and a white dwarf. The typical densities in the winds of the accretion disk and secondary are orders of magnitude too low to produce the observed X-rays, as is the mass of material in the accretion disk itself. The origin of the material is therefore unclear; the secondary could be an evolved object with a dense wind similar to that in some recurrent novae (in this case the implied mass-loss rate is $\sim 10^{-8}$ to $10^{-7} M_\odot \text{ yr}^{-1}$ for a wind speed of 20 km s^{-1}), or the material may have been ejected at low velocity at an early stage of the outburst. In either case, these observations of Nova Her 1991 should aid our understanding of the earliest stages of novae and the interaction of nova ejecta with their surroundings. $\square$

Received 9 September 1991; accepted 4 February 1991.

1. Bode, M. F. & Evans, A. (eds) *Classical Novae* (Wiley, Chichester, 1989).
2. Hoffman, J. A. *et al. Nature* **261**, 208–210 (1976).
3. Cruise, A. M. *Nature* **267**, 685–687 (1977).
4. Ögelman, H., Beuermann, K. & Krautter, J. *Astr. Astrophys.* **177**, 110–116 (1987).

5. Becker, R. H. in *Classical Novae* (eds Bode, M. F. & Evans, A.) (Wiley, Chichester, 1989).
6. Sogano, M. & Alcock, G. *IAU Circ. No. 5222* (1991).
7. Hjellming, R. J. *IAU Circ. No. 5234* (1991).
8. Wagner, R. M. *et al. IAU Circ. No. 5227* (1991).
9. Feast, M. W. *IAU Circ. No. 5230* (1991).
10. Feast, M. W. *IAU Circ. No. 5233* (1991).
11. Ashok, N. M. *et al. IAU Circ. No. 5254* (1991).
12. Trümper, J. *et al. Adv. Space. Res.* **2**, 241–249 (1983).
13. Predehl, P. *et al. IAU Circ. No. 5278* (1991).
14. Raymond, J. E. & Smith, B. W. *Astrophys. J. suppl. Ser.* **35**, 419–439 (1977).
15. Bath, G. T. & Harkness, R. P. in *Classical Novae* (eds Bode, M. F. & Evans, A.) (Wiley, Chichester, 1989).
16. Starrfield, S. in *Classical Novae* (eds Bode, M. F. & Evans, A.) (Wiley, Chichester, 1989).
17. Ögelman, H. *Lecture Not. Phys.* **369**, 148–154 (1990).
18. Warner, B. in *Classical Novae* (eds Bode, M. F. & Evans, A.) (Wiley, Chichester, 1989).
19. Gehrz, R. D. *et al. IAU Circ. No. 5236* (1991).
20. Bohlin, R. C., Savage, B. D. & Drake, J. F. *Astrophys. J.* **224**, 132–142 (1978).
21. Della Valle, M. *IAU Circ. No. 5223* (1991).
22. Granslo, B. H., Bortle, J. E., McNaught, R. H. & Hale, A. *IAU Circ. No. 5224* (1991).
23. Sonneborn, G., Yamamoto, M., Schmeer, P. & Royer, R. *IAU Circ. No. 5226* (1991).
24. Sveteck, P., McNaught, R. H. & Bortle, R. *IAU Circ. No. 5228* (1991).
25. McNaught, R. H., Schmeer, P. & Pereira, A. *IAU Circ. No. 5229* (1991).
26. Oshima, A. *IAU Circ. No. 5234* (1991).
27. Bortle, J. E., Schmeer, P., Hale, A., Korth, S. & Granslo, B. H. *IAU Circ. No. 5238* (1991).
28. Midtskogen, O., Granslo, B. H. & Mizser, A. *IAU Circ. No. 5265* (1991).
29. Synder, F. *IAU Circ. No. 5233* (1991).
30. Schmeer, P., Pearce, A., Halbach, E., Vendenet, M. & McKenna, J. *IAU Circ. No. 5254* (1991).

ACKNOWLEDGEMENTS. We thank D. Buczinski for informing us of the discovery of the nova and S. Starrfield for communicating his preliminary IUE results. H.M.L. is supported by the SERC.

Does an orbiting star cause periodic modulation of X-rays from NGC6814?

Marek Sikora* & Mitchell C. Begelman*

Joint Institute for Laboratory Astrophysics, University of Colorado and National Institute of Standards and Technology, Boulder, Colorado 80309, USA

USING data from the Ginga satellite, Done *et al.*¹ have confirmed an Exosat observation² of high-amplitude periodic modulation in the X-ray emission from the Seyfert galaxy NGC6814. To within observational errors of $\sim 1\%$, the period seems to have remained constant at $\sim 12,100$ seconds between 1985 and 1989. An obvious candidate for the phenomenon underlying the periodicity is the orbital motion of a star or low-mass compact object around the central black hole³. As we show here, the presence of an orbiting star could be verified easily by looking for the effects of Lense-Thirring precession of the orbital plane, caused by the dragging of inertial frames around a rotating black hole. Precession-induced variations in the waveform and in the phase of the observed periodicity should have a period of between a month and a year. Such variations could account for the different waveforms present in the Ginga and Exosat data sets⁴, and may be detectable in existing Ginga and future Rosat, OSSE/GRO, and Astro-D data.

Let us assume that a star of mass $m \ll M_{\text{BH}}$ orbits the central black hole in NGC6814. A value for the black-hole mass of the order $10^6 M_{\odot}$ is suggested by the time lag between variations in the hard X-ray flux and the iron $K\alpha$ fluorescence line⁵ and the relatively low bolometric luminosity of NGC6814 ($\leq 10^{44} \text{ erg s}^{-1}$, close to the Eddington limit for a $10^6 M_{\odot}$ object). The radius of the orbit (supposed to be circular³), in units of the gravitational radius $r_g \equiv GM_{\text{BH}}/c^2 = 1.5 \times 10^{11} M_6 \text{ cm}$, is $x \equiv r/r_g = 54 \tilde{P}^{2/3} M_6^{-2/3}$, where $\tilde{P} \equiv 1$ is the orbital period normalized to 12,100 s (included to show the dependence) and $M_6 = M_{\text{BH}}/10^6 M_{\odot}$. We conjecture that the periodic modulations are due to a misalignment between the plane of the orbit and the plane of an accretion disk. Highly eccentric stellar orbits around the central black hole will be made more circular by interaction

with the disk before they are brought into alignment with the disk plane³. At such small distances from the black hole, the angular momentum vector of the accretion disk is likely to be aligned with the spin axis of the central black hole⁶, whereas the inclined stellar orbit will precess about this axis as a result of the dragging of inertial frames by the rotating black hole. This effect was discovered by Lense and Thirring in 1918⁷ in the weak field limit, and derived for the Kerr metric, which describes the gravitational field of a rotating black hole, by Wilkins⁸. The precession period is

$$P_{\text{prec}} = \frac{\pi r_g}{c} \left(\frac{J}{J_{\text{max}}} \right)^{-1} x^3 = 2.4 \times 10^6 \tilde{P}^2 M_6^{-1} \left(\frac{J}{J_{\text{max}}} \right)^{-1} \text{ s}$$

where (J/J_{max}) is the ratio of the black hole's angular momentum to the maximum possible value. If the black hole grew by accreting material with high angular momentum, then we expect (J/J_{max}) to be of order unity. As M_6 is unlikely to be much smaller than one (to avoid violating the Eddington limit), we conclude that the precession period must be of order a year or less.

The gravitational radiation timescale places the strongest upper limit on the mass of the body orbiting the central black hole:

$$\left| \frac{1}{P} \frac{dP}{dt} \right|_{\text{gr}}^{-1} = 6.6 \times 10^{-2} \tilde{P}^{8/3} M_6^{-5/3} \left(\frac{m}{M_{\text{BH}}} \right)^{-1} \text{ yr}$$

As the period has apparently changed by no more than 1% between 1985 and 1989⁴, the mass of the companion cannot exceed $100 M_6^{-2/3} M_{\odot}$. If the companion is a normal star, we can obtain a constraint on its radius from the tidal limit:

$$r_* < x r_g \left(\frac{m}{M_{\text{BH}}} \right)^{1/3} = 8.1 \times 10^{10} \tilde{P}^{2/3} \left(\frac{m}{1 M_{\odot}} \right)^{1/3} \text{ cm}$$

Solar-type stars would probably suffer considerable damage from tidal stresses, and much more massive main-sequence stars and giants would be completely disrupted. To resist the tides, the star would have to be a dwarf somewhat smaller than the Sun, or a compact object.

Each time the star crosses the accretion disk, it shocks a plug of material which subsequently expands away from the disk, trailing in the wake of the star. Although the star collides with the disk twice per orbit, the expected large optical depth of the disk leads us to associate the reported modulations with the star's crossing from the far side to the near side. There are at least three ways in which this process might lead to periodic modulation of the X-ray output: the orbital kinetic energy dissipated in the collision between the star and the accretion disk could power an X-ray flare; the interaction of the ejected matter with a centrally produced wind or jet could lead to enhanced X-ray emission³; and the ejected matter could occult the central X-ray source. Although it is not our main purpose here to present a detailed model for the X-ray modulations, we comment briefly on each of these possibilities below.

The shocked disk material itself is not expected to be a significant source of X-rays³. The optical depth of the disk is so large that virtually all of the dissipated energy goes into driving the expansion of the shocked gas. But an X-ray flare is possible if the expanding gas transfers its kinetic energy to an optically thin (but sufficiently dense) accretion disk corona, either by shock-heating it or by triggering widespread reconnection in the embedded magnetic field. The total energy dissipated per star-disk encounter is roughly $\Delta E \approx \frac{1}{2} \Delta m g^2 v_*^2$, where Δm is the swept-up mass, $v_* = 40,000 M_6^{1/3} \text{ km s}^{-1}$ is the keplerian velocity of the star, and $g = 2^{1/2} (1 - \cos i)^{1/2}$, where i is the inclination angle of the stellar orbit relative to the disk. To dissipate 10^{46} erg , roughly 10 times the X-ray energy contained in a typical modulation², the star would have to sweep up $\sim 6 \times 10^{-7} g^{-2} M_{\odot}$. For a star interacting across its geometric cross-section, this corresponds to a disk column density

*M.S. is on leave from N. Copernicus Astronomical Center, Warsaw, Poland. M.C.B. is also at Department of Astrophysical, Planetary and Atmospheric Sciences, University of Colorado, Boulder, Colorado, USA.

$\geq 10^{28} g^{-2} (r_*/r_\odot)^{-2} \text{ cm}^{-2}$ and a Thomson optical depth $\geq 10^4 g^{-2} (r_*/r_\odot)^{-2}$. Column densities of this order are predicted by a standard accretion-disk model⁹ for NGC6814, provided that the viscosity parameter α is 10^{-2} or smaller, i is not $\ll 1$, and the star is not too much smaller than the Sun. Note that the high values of Δm required imply that the stellar orbit will become aligned with the disk on a relatively short timescale. According to calculations³, the alignment timescale will be $\leq 200 i^2 (m/1 M_\odot) \text{ yr}$ for $i \ll 1$, increasing to as much as $10^3 (m/1 M_\odot) \text{ yr}$ for orbits inclined at $\sim 45^\circ$. Thus, the energetics are marginally self-consistent, provided that the orbital inclination is not too small.

The second case, in which flares are induced by interaction of the ejected matter with a central outflow, faces problems with the flare amplitude in the case of a quasispherical wind, because of the low covering factor of the ejected matter. If the interaction occurs with a highly collimated jet, a finely tuned orbital alignment is required with $i \sim \pi/2$.

The third possibility, occultation of the central source by the ejected matter, provides the easiest way to explain the existence of periodic X-ray modulations. The constraints on the column density of the disk, the orbital inclination and the radius of the star are not as severe as in the first case. If the main body of the ejected and expanding matter is sufficiently opaque for Thomson scattering, then, for certain locations of the observer, it will impose dips on the intrinsically variable flux produced in the central source. Radiation from the central source can also be modulated by the partially opaque 'tail' of the ejected gas, as well as by matter overflowing the Roche lobe of the star. The latter may be especially important if the star is 'bloated' as a result of the strong irradiation close to the central engine¹⁰⁻¹². We note that 'dips and sharp drops' are observed in at least two of the three Ginga data sets¹³, with the X-ray flux falling to very low or negligible levels on timescales as short as 50 s. These features suggest that the central X-ray source is occulted by optically thick material with a sharp 'edge'.

Fortunately, our main conclusions are independent of the details of how the X-rays are modulated. In any model relating the periodicity in NGC6814 to the keplerian period of a star on an inclined orbit around the black hole, the plane of the star's orbit must undergo Lense-Thirring precession with a period of a month to a year. Because the appearance of a flare or dip is likely to depend strongly on geometry, we suggest that the characteristic waveform may vary on a similar timescale. These variations may explain the different waveforms observed by Exosat in 1985 and Ginga in 1989. Periodic variations on the precession timescale may also be responsible for differences in the characters of sharp dips detected during the first two Ginga observations on NGC6814, and the lack of such dips during the third observation⁴. These observations each lasted 2-3 days, and were separated by 1 year and 6 months, respectively. Variations caused by precession of the orbit may also show up in the properties of the iron fluorescence line and the absorption due to the 'warm absorber', as we expect a considerable amount of ejected material to linger above the disk. But the most unambiguous prediction is that the orbital phase at which the collision between star and accretion disk occurs will appear to oscillate periodically with the precession frequency. Because of the light travel time across the stellar orbit, star-disk collisions that occur on the far (near) side of the black hole will appear to lag (lead) the true orbital phase of the collision by $270 M_6^{1/3} \cos \theta_{\text{obs}}$ s, or as much as 2% of the orbital period, where θ_{obs} is the inclination of the line of sight relative to the disk. This will lead to periodic shifts in the phase at which the X-ray modulation occurs. As the modulation may not be in phase with the collision (especially in the case of an occultation model, where material ejected on the far side has to be carried around to the near side at the keplerian speed before it occults the centre) the observed phase lag in the modulation may be as large as 50%. A thorough study of these phase shifts would help to determine the mechanism

responsible for the modulation. Detection of the precession would constitute direct support for models of active galactic nuclei involving disk accretion onto a supermassive black hole. $\square$

Received 11 November 1991; accepted 7 February 1992.

1. Done, C. et al. in *Structure and Emission Properties of Accretion Disks* (eds Bertout, C., Collin-Souffrin, S., Lasota, J.-P. & Tran Thanh Van, J.) 417-418 (Editions Frontières, Gif sur Yvette, 1991).
2. Mittaz, J. P. D. & Branduardi-Raymont, G. *Mon. Not. R. astr. Soc.* **238**, 1029-1046 (1989).
3. Syer, D., Clarke, C. J. & Rees, M. J. *Mon. Not. R. astr. Soc.* **250**, 505-512 (1991).
4. Done, C. et al. in *Testing the AGN Paradigm* (eds Urry, C. M., Neff, S. & Holt, S.) (AIP Conf. Proc. Series, in the press).
5. Done, C. et al. in *Frontiers of X-ray Astronomy* (eds Tanaka, Y. & Koyama, K.) (Universal Academy Press, Tokyo, in the press).
6. Bardeen, J. M. & Petterson, J. A. *Astrophys. J.* **195**, L65-67 (1975).
7. Lense, J. & Thirring, H. *Physik. Z.* **19**, 156-163 (1918).
8. Wilkins, D. C. *Phys. Rev. D* **5**, 814-822 (1972).
9. Shakura, N. I. & Sunyaev, R. A. *Astr. Astrophys.* **24**, 337-355 (1973).
10. Tout, C. A., Eggleton, P. P., Fabian, A. C. & Pringle, J. E. *Mon. Not. R. astr. Soc.* **238**, 427-438 (1989).
11. MacDonald, J., Stanev, T. & Biermann, P. L. *Astrophys. J.* **378**, 30-38 (1991).
12. Czerny, M., Sikora, M. & Begelman, M. C. in *Relativistic Hadrons in Cosmic Compact Objects* (eds Zdziarski, A. A. & Sikora, M.) 23-25 (Springer, Heidelberg, 1991).
13. Kunieda, H. et al. in *Iron Line Diagnostics in X-ray Sources* (ed. Treves, A.) (Springer, Berlin, in the press).

ACKNOWLEDGEMENTS. We thank C. Done for discussions and for comments on the manuscript. M.S. thanks the High Energy Astrophysics group at NASA/Goddard Space Flight Center for their hospitality. This research was supported by the NSF and NASA.

Optical limiting performance of C_{60} and C_{70} solutions

Lee W. Tutt & Alan Kost

Hughes Research Laboratories, Malibu, California 90265, USA

OPTICAL sensors used in connection with bright sources such as lasers and arc welders must commonly be protected from damaging light levels by the use of optical limiters¹. One approach to optical limiting makes use of materials whose optical transmittance decreases at high light levels²⁻⁵. For most protective applications, the response must be rapid and the saturation threshold low; a lower threshold provides a greater safety margin. Studies of the optical properties of C_{60} have shown that the absorption cross-section of the photoexcited triplet state is greater than that of the ground state⁶, suggesting that it may have a nonlinear optical response of the sort useful for optical limiting. Here we report measurements of the optical response of solutions of C_{60} and C_{70} in methylene chloride and toluene, using 8-ns pulses of 532-nm-wavelength laser light. We observed optical limiting behaviour in all cases, with saturation thresholds equal to or lower than those reported for other optical-limiting materials currently in use.

Electro-evaporated soot was purchased from the MER Corporation. The toluene-soluble portion of the soot was separated and purified by the method of ref. 7. The measuring apparatus has previously been described². We use temporally smooth, spatially uniform, optical pulses from a mode-locked, frequency-doubled Nd-YAG laser as the excitation source, and measure the incident and transmitted energy for each pulse. The output from the laser is collimated. The incident fluence is first increased and then decreased to check for evidence of photodegradation.

The optical limiting responses of a 63% and an 80% transmitting toluene solution of C_{60} to 532-nm optical pulses are shown in Fig. 1. At very low fluences the optical response of the solution obeys Beer's law and the transmittance is roughly constant. At an incident intensity of $\sim 100 \text{ mJ cm}^{-2}$ the transmittance begins to decrease markedly. The transmitted fluence effectively becomes clamped at $\sim 65 \text{ mJ cm}^{-2}$. The 80% transmitting sample behaved similarly, clamping at 240 mJ out to greater than 3 J cm^{-2} . Solutions of C_{60} in methylene chloride give similar results indicating solvent effects are small.

Figure 2 is a comparison of the performance of C_{60} to that which has been reported for other soluble, absorbing materials

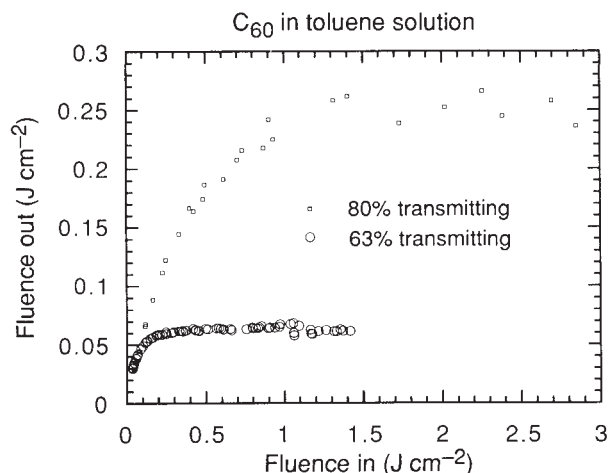


FIG. 1 Optical limiting response of 63% and 80% transmitting solutions of C_{60} in toluene to 7 ns, 532 nm optical pulses.

whose transmittance saturates at a new lower level ('reverse-saturable' materials)²⁻⁵. We measured the performance of each of these compounds as a 70% transmitting solution at 532 nm. The optical limiter with the lowest threshold for induced absorption is C_{60} . It should be noted that chloroaluminum phthalocyanine is a limiter at 532 nm but is a saturable absorber at 650 nm (refs 7, 8).

The optical limiting properties are consistent with the efficient population of a triplet state with a much higher absorption cross-section than the ground state. For pulse lengths less than the lifetime of the triplet state, the triplet state will act as an accumulation site. The limiting action of C_{60} solutions will be most effective for pulses less than 40 μ s, the triplet state lifetime. For a longer pulse, a considerable fraction of the excited molecules decay back to the ground state before the end of the pulse. The triplet-triplet absorption spectrum roughly follows the ground-state absorption⁶; we therefore expect the limiting action to be very broadband. Although there may be a nonlinear scattering component to the limiting action, the dominant mechanism is reverse saturable absorption as evidenced by the small solvent dependence.

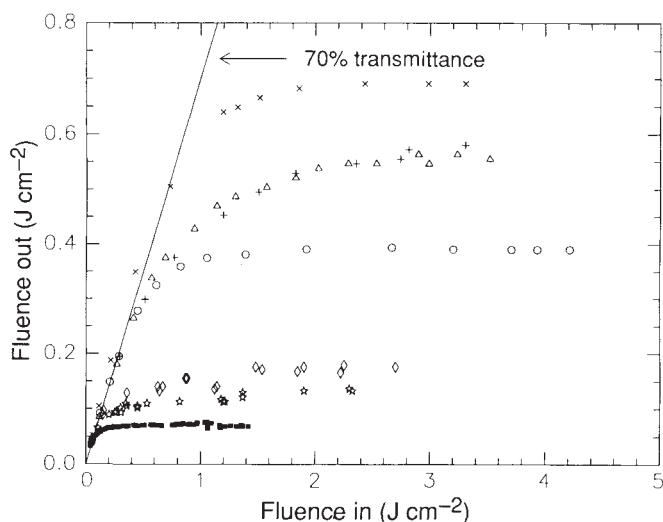


FIG. 2 Comparison of the optical limiting response of solutions of various reported optical limiters to C_{60} in toluene. All comparison solutions are 70% transmitting at 532 nm, and the solvent was methylene chloride, except with the two exceptions of chloroaluminum phthalocyanine, which was dissolved in methanol, and indanthrone, dissolved in dilute KOH. $\circ$, $HfFeCo_3(CO)_{10}$ ($(P(CH_3)_3)_2$); Δ , $HfFeCo_3(CO)_{12}$; +, $(N(C_2H_5)_4)^+(FeCo_3(CO)_{12})^-$; $\times$, $HfFeCo_3(CO)_{10}$ ($(PPh_3)_2$); $\diamond$, indanthrone; $\star$, chloroaluminum phthalocyanine; $\blacksquare$, C_{60} .

Solutions of C_{70} also show optical limiting action. The output limiting fluence is $\sim 350 \text{ mJ cm}^{-2}$ for a 70% C_{70} solution in toluene. The reddish-orange C_{70} solutions have a higher absorption at 532 nm than the magenta C_{60} solutions. It is therefore not surprising that the higher ground-state absorption of C_{70} results in a smaller ratio of excited-state to ground-state absorption and a higher threshold for optical limiting.

For many optical limiting applications, solid films are desirable. Further work is underway to elucidate the response of fullerene films to optical pulses. $\square$

Received 13 November 1991; accepted 11 February 1992.

1. Wood, G. L., Clark, W. W. III, Miller, M. J., Salamo, G. J. & Sharp, E. J. *SPIE Proc.* **1105**, 154-180 (1980).
2. Tutt, L. W. & McCahon, S. W. *Opt. Lett.* **15**, 700-703 (1990).
3. Hoffman, R. C., Stetzyk, K. A., Potember, R. S. & McLean, D. G. *J. opt. Soc. Am.* **B6**, 772-777 (1989).
4. Tutt, L. W., McCahon, S. W., & Klein, M. B. *SPIE Proc.* **1307**, 315-326 (1990).
5. Coulter, D. R. *et al.* *SPIE Proc.* **1105**, 42-51 (1989).
6. Arbogast, J. W. *et al.* *J. phys. Chem.* **95**, 11-12 (1991).
7. Ajie, H. *et al.* *J. phys. Chem.* **94**, 8630-8633 (1990).
8. Lund, D. J., Edsall, P. & Masso, J. D. *SPIE Proc.* **1207**, 193-201 (1990).

ACKNOWLEDGEMENTS. We thank S. W. McCahon and M. B. Klein for valuable discussions.

Young's modulus of a solid two-dimensional Langmuir monolayer

H. Bercegol* & J. Meunier

Laboratoire de Physique Statistique de l'ENS, 24 rue Lhomond, 75231 Paris Cedex 05, France

LANGMUIR monolayers—films of amphiphilic molecules at the surface of water—exhibit many phases^{1,2}. Some of these behave like two-dimensional solids on experimental timescales, but previous measurements of the shear modulus of these 'solid' monolayers³⁻⁵ have yielded a value too small to be compatible with a two-dimensional crystal. The interpretation of these is complicated, however, by the likelihood of inhomogeneities in the films, which are probably assemblies of microscopic crystalline domains. Here we describe measurements of the Young's modulus of an isolated 'solid' domain of NBD-stearic acid monolayers. We obtain a value large enough to be compatible with the modulus of a two-dimensional crystal⁶⁻⁸. This suggests that Langmuir monolayers should provide model systems for studies of melting in two dimensions⁶⁻⁸.

The compressional modulus, K , of a Langmuir monolayer is the derivative of surface pressure, Π , with respect to the change in area per molecule, and is large in the condensed phase: $K \approx 1,000 \text{ mN m}^{-1}$. Mouquin and Rideal⁹ introduced a method of measuring the shear modulus μ by applying a torque on the monolayer. Results have recently been obtained for stearic acid³ and alcohols^{4,5}. Measured values of μ , except in the presence of trivalent ions (Fe^{3+} or Al^{3+}), are very small (in the range 10^{-2} to 10 mN m^{-1}). The two-dimensional (2D) Young's modulus E_{2D} deduced from these measurements are thus very small: $E_{2D} = 4\mu K / (\mu + K) \approx 4\mu$. Theoretical predictions developed^{6,8} from the initial ideas of Kosterlitz-Thouless (melting by dislocation unbinding) give a lower limit for the Young's modulus of a two-dimensional crystalline solid of $E_{min} = 16\pi k_B T_f / \Sigma_s$, where k_B is the Boltzmann constant and Σ_s the area per molecule in the solid phase at the melting point T_f . With $T_f = 300 \text{ K}$ and $\Sigma_s = 33 \text{ \AA}^2$ per molecule, one obtains $E_{min} \approx 630 \text{ mN m}^{-1}$, a value much larger than the experimental Young's modulus deduced from shear experiments³⁻⁵. This seems to prove that the dense phases in Langmuir monolayers cannot be true crystalline solid phases. But the measurements are made on macroscopic

* Present address: Department of Materials Science and Engineering, University of Cincinnati, Mail location 12, Cincinnati, Ohio 45221, USA.

monolayers, which are probably not homogeneous. If so, the measured modulus is not that of the 'solid' phase, but that of an aggregate of small domains of 'solid' phase, which may contain liquid¹⁰ (Fig. 1a). To avoid these difficulties, we have directly measured the Young's modulus of a single domain of the 'solid' phase.

The amphiphilic molecule studied is 12-NBD-stearic acid (Molecular Probes, Eugene, Oregon, USA). This molecule fluoresces when illuminated with blue light, allowing us to observe phase transitions by fluorescence microscopy without adding any fluorescent impurity¹⁰. The Langmuir monolayer obtained with this molecule is homogeneous and liquid if Σ is in the range 84 to 110 Å² molecule⁻¹ at $T = 20^\circ\text{C}$. For $\Sigma < 84$ Å² molecule⁻¹, there is a first-order phase transition, and 'solid' domains appear in the liquid phase. In this 'solid' phase the area per molecule is 33 Å² molecule⁻¹ or less. The first reports indicated that the domains were needle-shaped¹⁰, but in the absence of impurities due to oxidation of the molecule at room temperature, the 'solid' domains look like long rods with two parallel sides¹¹ (Fig. 1b). Rods directly obtained by depositing the pure product at the free surface of water, without solvent, have a length L of a few millimetres, and a width e of tens of micrometres. This domain shape is well adapted to a direct measurement of the longitudinal Young's modulus through a micromechanical experiment. The flexion s_f of a rod is measured

when a force f is applied at its mid-point while its ends are supported (f in the plane of the water and perpendicular to the length of the rod). The Young's modulus is then $E_{2D} = fL^3/48s_fI$, where $I = e^3/12$ is the flexional rigidity.

To apply the force on the domain centre and support the ends, we use three glass fibres ($\sim 10\ \mu\text{m}$ in diameter), fixed at one end and perpendicular to the surface of water. The two outer fibres, separated by ~ 1 mm, are rigid because short, their fixing point being ~ 2 mm below the surface. The third is more flexible, as its fixing point is ~ 1 cm above the surface. A micrometer screw moves this fibre along the perpendicular bisector of the line joining the points where the two rigid fibres cross the surface of the water. A rod is introduced between the three fibres, and a force is applied at its centre by moving the flexible fibre (Fig. 2a). The difference between the displacement of the fixing point of the flexible fibre (measured with the micrometer screw) and the displacement of the point where it crosses the interface (measured on the microscope images) gives the flexion s_f of the fibre at the surface and consequently the applied force on the rod: $f = \kappa s_f$ (with $\kappa = 0.16\ \text{mN m}^{-1}$).

Three forces act in addition to the elastic force of a fibre. The first is the gravity which acts on each element of the fibre. The component of the weight normal to the fibre bends it. This is equivalent to a (negligible) horizontal force in the plane of the water: $f_p \ll \kappa_p s_f$ ($\kappa_p \approx 2 \times 10^{-3}\ \text{mN m}^{-1}$), much smaller than f .

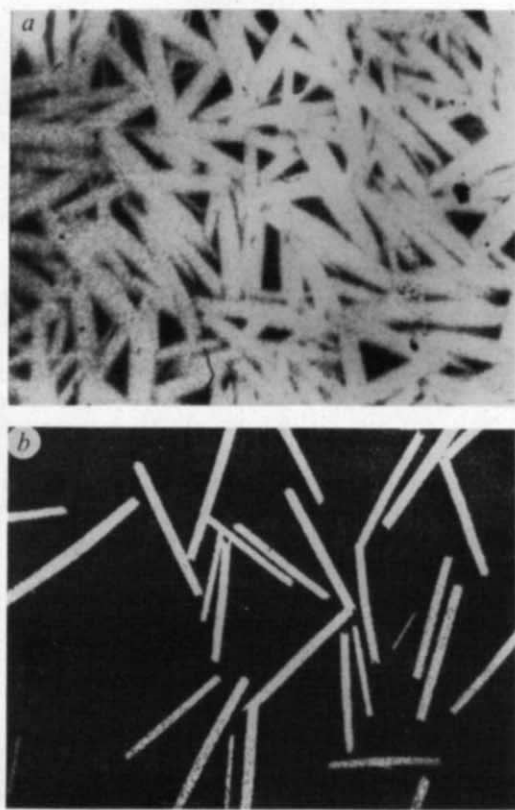


FIG. 1 *a*, Monolayer of fluorescent amphiphilic molecules, 12-NBD-stearic acid, after compression from the fluid phase into the 'solid' region¹⁰. The mean molecular area is ~ 30 Å² molecule⁻¹. The monolayer is a collection of microdomains of a fluid (dark) and a solid (bright) phase. Further compression steeply increases the surface pressure but does not make the fluid region disappear: the monolayer collapses before the disappearance of inhomogeneity. The area shown is $320 \times 240\ \mu\text{m}^2$. *b*, Solid domains of pure 12-NBD-stearic acid are rectangular rods. These were obtained by compression of a monolayer deposited from solution in chloroform. Mean molecular area is 65 Å² molecule⁻¹. Some slightly bent domain edges seen here are probably caused by residual oxidized impurities (as in ref. 11). Scale as in *a*.

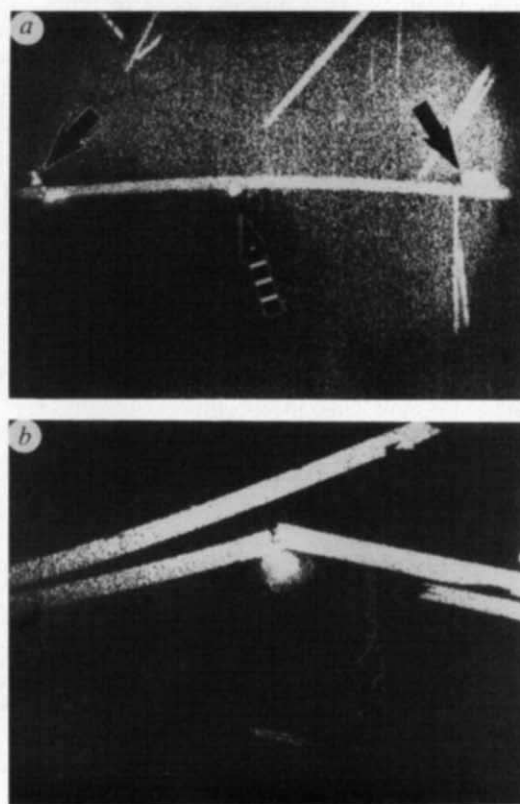


FIG. 2 *a*, Bending a two-dimensional rod with three glass fibres. The fibres are perpendicular to the page; they are visible because of the deposition of fluorescent amphiphiles on their surface. Two rigid fibres (shown by black arrows) are separated by 1.1 mm. The flexible fibre (dashed arrow) exerts a force on the middle of the rod and bends it. The domains were obtained by compression of a monolayer deposited from the three-dimensional gel-like state of the amphiphile, without any solvent: this process gives bigger domains than those of Fig. 2. Area shown is $1,300 \times 950\ \mu\text{m}^2$. *b*, A rod broken by the action of the fibre. The fibre is visible in the middle of the photograph, near the break. Area shown is $640 \times 480\ \mu\text{m}^2$.

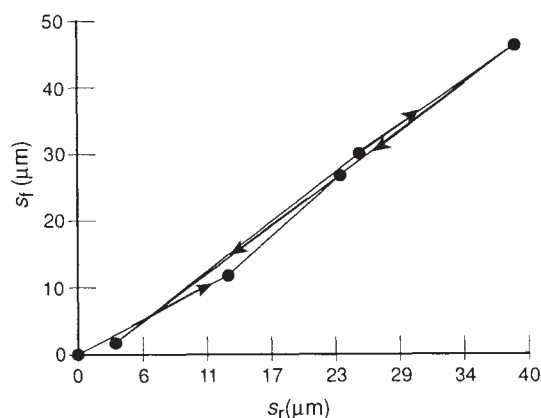


FIG. 3 Reversible flexion experiment. The flexion s_f of the glass fibre depends linearly on the flexion s_r of the rod. Arrows indicate the sense of capture of experimental points; each measurement is taken at static equilibrium. Width of the rod is $\sim 25 \mu\text{m}$.

The two remaining forces are the capillary forces and the hydrostatic pressure. The capillary rise h along a fibre is small, of the order of the diameter ϕ of the fibre: $h \approx 3\phi \cos \theta$, where θ is the wetting angle on the fibre. Consequently, the ratio between the hydrostatic force acting on one side of a fibre and that due to the component of the capillary forces normal to the same side of the fibre, $\Delta f_{\text{pressure}}/\Delta f_{\text{tension}} = \rho g(3\phi \cos \theta)^2/2\gamma \sin \theta \approx 10^{-5} \cos^2 \theta/\sin \theta$ (where γ is the interfacial tension), is small unless θ is close to 0 or π . For a constant contact angle, the sum of the capillary forces that act on a vertical and isolated fibre is directed along the fibre and does not produce any bending. This is not the case when the fibre is inclined because of bending. An estimation of the capillary force normal to the fibre gives $f_i \approx \kappa_i \cos \theta s_r$ ($\kappa_i \approx 10^{-1} \text{ mN m}^{-1}$). Another effect of the capillarity is to produce an attraction between two fibres. An estimate of this force¹² gives: $df_{\text{inter}}/dr \approx \pi/2 (\phi/r)^2 \gamma \cos \theta \sin \theta \approx 3 \times 10^{-2} \cos \theta \sin \theta (\text{mN m}^{-1})$ where r is the distance between the fibres ($\sim 0.5 \text{ mm}$). Capillary and hydrostatic effects are negligible as soon as $\theta \approx \pi/2$. We therefore tried to obtain a wetting angle θ as close to $\pi/2$ as possible. In these conditions the capillary rise h is very small. This has been obtained in our experimental conditions (water covered with a dense monolayer) by using silanized glass fibres.

A difference of contact angle between the sides of the fibre introduces a force $f_{\Delta\theta} \approx \Delta(\gamma\phi \sin \theta)_{\theta} \approx 6 \times 10^{-4} \cos \theta \Delta\theta (\text{mN})$, where $\Delta\theta$ is the difference in this angle. This force can be of the same order of magnitude as f but is important in our experiment only if it changes (that is, the contact angle varies)

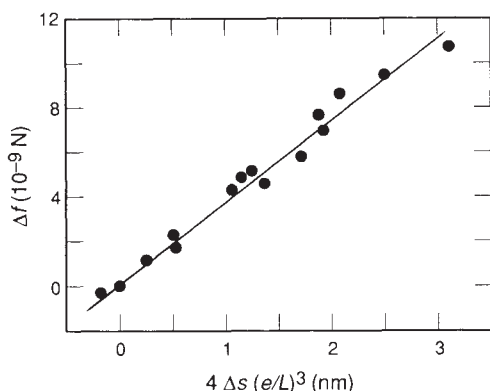


FIG. 4 Results of five reversible flexion experiments with five different rods. Rod widths range from 17 to $30 \mu\text{m}$. The zero of each experiment has been redefined because it is not known exactly. The plotted quantities are therefore a difference of forces, Δf , against a difference in rod flexion, Δs . This Δs is multiplied by a geometrical factor accounting for difference in rod width, e , and interval between rigid fibres L . The slope of this linear curve gives the Young's modulus of the two-dimensional solid.

during the measurement. This happens when amphiphilic molecules form a deposit on the fibre, an irreversible transformation.

The force f applied on a rod was plotted against its flexion s_r . Results are shown only for reversible experiments giving the same result when the force is increased or decreased (Fig. 3). An irreversible experiment indicates that the contact angle changed during the measurement. The elimination of such experiments is justified because the effect of contact angle change is random; the measured modulus is randomly augmented or diminished by this transformation. The Young's modulus E deduced from a plot of the applied force against $4s_r(e/L)^3$ for several rods (Fig. 4) $E_{2D} = 3,660 \text{ mN m}^{-1}$. The dispersion of the measurements is small, 10% (Fig. 4) but the accuracy of the measurement is low ($\pm 1,300 \text{ mN m}^{-1}$), mainly because a systematic error is introduced by errors in measuring the diameter of the fibres (5%), their length (1%) and the Young's modulus of the glass (1%) to determine the constant κ .

The E_{2D} value obtained is compatible with the minimum value E_{min} for a two-dimensional crystalline solid, but it must be remarked that the Kosterlitz-Thouless theory was developed for a solid with isotropic mechanical properties. This is probably not the case for the rods we have studied, which are optically anisotropic¹⁰. Nevertheless, one does not expect important changes in the predicted value when anisotropy is introduced¹³.

No plastic deformation of the rods was observed when the constraint is relaxed. When the constraint is large, the rods break in two without plastic deformation (Fig. 2b). We have not yet studied the breaking systematically (experiments are in progress) but we observed that if there is a threshold of breaking, it is for a strain smaller than a few 10^{-3} .

The measured Young's modulus E_{2D} is that of a two-dimensional solid. An estimate of the thickness of the monolayer gives $l = 18.5 \text{ \AA}$. The corresponding three-dimensional modulus is $E_{3D} = E_{2D}/l \approx 2 \times 10^9 \text{ Pa}$, which is 70 times smaller than that of iron and 35 times smaller than that of glass, but 1,000 larger than that of rubber. □

Received 20 December 1991; accepted 30 January 1992.

1. Qiu, X., Ruiz-Garcia, J., Stine, K. J., Knobler, C. M. & Selinger, J. V. *Phys. Rev. Lett.* **67**, 703-706 (1991).
2. Bibo, A. M. & Peterson, I. R. *Adv. Mat.* **2**, 309-311 (1990).
3. Abraham, B. M., Ketterson, J. B., Miyano, K. & Kueny, A. *J. chem. Phys.* **75**, 3137-3141 (1981).
4. Abraham, B. M., Miyano, K., Xu, S. Q. & Ketterson, J. B. *Phys. Rev. Lett.* **49**, 1643-1646 (1982).
5. Abraham, B. M. & Miyano, K. *Phys. Rev. Lett.* **51**, 1975-1978 (1983).
6. Nelson, D. R. & Halperin, B. I. *Phys. Rev.* **B19**, 2457-2484 (1979).
7. Young, A. P. *Phys. Rev.* **B19**, 1855-1866 (1979).
8. Nelson, D. R. in *Phase Transitions and Critical Phenomena*, Vol. 7 (eds Domb, C. & Lebowitz, J. L.) 2-93 (Academic, London, 1983).
9. Mouquin, H. & Rideal, E. K. *Proc. R. Soc. A* **114**, 690-697 (1927).
10. Bercegol, H., Gallet, F., Langevin, D. & Meunier, J. *J. Phys.* **50**, 2277-2289 (1989).
11. Bercegol, H. *J. phys. Chem.* (in the press).
12. Cloitre M. thesis, Univ. of Paris VI (1989).
13. Ostlund, S. & Halperin, B. I. *Phys. Rev.* **B23**, 335-358 (1981).

ACKNOWLEDGEMENTS. H.B. thanks Y. Pomeau for discussions and C. Lemaire for technical help. The E-glass fibres were provided by D. Frugier of Saint-Gobain Recherche.

Nutrient control of phytoplankton photosynthesis in the Western North Atlantic

Trevor Platt*, Shubha Sathyendranath*†, Osvaldo Ulloa†, William G. Harrison*, Nicolas Hoepffner† & Joaquim Goes‡

* Biological Oceanography Division, Bedford Institute of Oceanography, Box 1006, Dartmouth, Nova Scotia B2Y 4A2, Canada

† Department of Oceanography, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

‡ Biological Oceanography Division, National Institute of Oceanography, Dona Paula, Goa 403 004, India

LIMITED understanding of the influence of the environment on the relation between photosynthesis and light in the ocean impairs our capacity to estimate primary production from remotely sensed data on ocean colour¹ and to model the role of the marine biota in the ocean carbon cycle². Here we report results from several years of oceanographic cruises, showing that the parameters of the photosynthesis–light curve for the flora of the North Sargasso Sea are remarkably constant in magnitude, except during the spring phytoplankton bloom when their magnitudes are noticeably higher. We interpret these results as providing direct evidence for nutrient control of photosynthesis in the open ocean. Our findings also reinforce the plausibility of using biogeochemical provinces to partition the ocean into manageable units for basin- or global-scale analysis^{1,3,4}, show that seasonal changes in critical parameters should not be overlooked if robust carbon budgets are to be constructed, and illustrate the value of attacking the parameters that control the key fluxes, rather than the fluxes themselves, when investigating the ocean carbon cycle.

The dependence of phytoplankton photosynthesis on light can be determined routinely during oceanographic cruises. Each light-saturation curve is described by two parameters: the initial slope α^B , an index of the efficiency of photosynthesis in weak light, and the assimilation number P_m^B , a measure of the rate of photosynthesis in saturating light⁵, where the superscripts indicate normalization to the biomass B . At any time and place, knowledge of these two parameters is sufficient to calculate primary production, given information on light and biomass. Here we report the results of ~152 characterizations of the light-saturation curve collected during six cruises to the North Sargasso Sea (25–40° N, 40–70° W). The 1989 (North Atlantic pilot study) and 1990 cruises represent Canadian participation in the Joint Global Ocean Flux Study.

The average magnitude of α^B was the same for each of the first four cruises (Table 1), covering, in four different years, a range of conditions from the immediate post-spring bloom (nitrate still measurable in the upper 50 m) through the oligotrophic summer phase (nitrate undetectable at most of the stations in the upper 50 m) to the presumed end of a late summer/early autumn phytoplankton bloom (presumption based on near-surface nitrate levels).

In the spring bloom conditions of the fifth cruise (1989), however, the mean value of α^B was more than twice as high as for the other four cruises. The corresponding increases in nitrate and chlorophyll were about a factor of 10. The 1990 cruise covered the transition between the spring bloom and the summer oligotrophic phase, such that the mean α^B for the cruise is intermediate between that of the first four cruises and that of the 1989 cruise. These 1990 data, collected during the declining phase of the spring bloom, show a clear time dependence, with a steady decrease in α^B from the start to the end of the cruise (Fig. 1a), the ship staying in the same parcel of water as tracked by a free-floating buoy. Nitrate data for this period indicate that spring bloom conditions prevailed initially, with surface values

near 1.5 mg-at N m⁻³, decreasing steadily with time to below the level of detection at the end of the cruise.

Field data that show direct evidence for nutrient control of primary production are rare for open ocean conditions, lending added interest to these observations (Fig. 1b). The results suggest that we should account for the effect of environmental covariables, such as nutrients, through their influence on the relevant parameters rather than on the fluxes themselves. Comparison of the vertical structure of α^B across all the cruises (Fig. 2) also reveals the influence of nutrients, although in this case it is complicated by the effects of photoadaptation. The summer months (June–September), after the development of density stratification, are characterized by values of α^B which are very low in the top 70 m (average ~0.037 mg C (mg chlorophyll)⁻¹ h⁻¹ (W m⁻²)⁻¹), but which increase rapidly with further increase in depth (mean values about twice as high as in the surface layer). In our data, this effect is most pronounced in September (1988), when the stratification is strongest. On the other hand, data collected in Spring show no clear depth dependence. These results are readily understandable as a consequence of the opposition between photoadaptation and vertical mixing. In a stratified water column, cells in the deeper layer tend to be confined to that layer long enough for them to adapt to its low average light level^{6,7}; moreover, cells in the deeper layer retain access to the deep nutrient reservoirs, whereas the upper layer is typically depleted of nutrients.

The assimilation number P_m^B was also higher by a factor of 2–3 during the spring bloom compared with the summer oligotrophic phase (Table 1). Like α^B , the assimilation number was stratified by depth in the summer months, with cruise averages for the top 70 m in the range from 1.8 to 3.4 mg C (mg

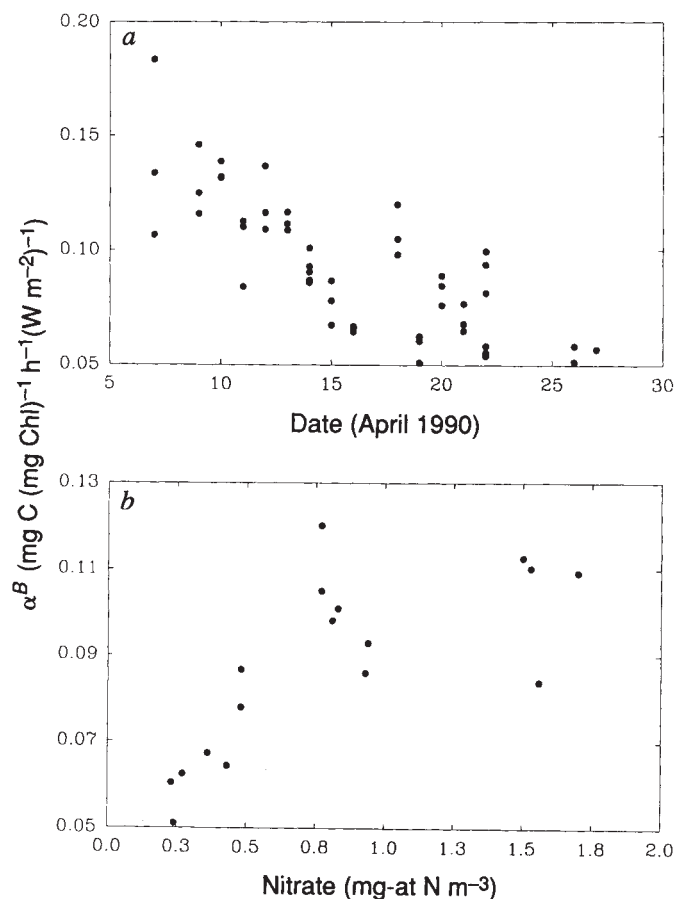


FIG. 1 a, Time dependence of α^B , for the April 1990 cruise. b, Variation of α^B with nitrate for samples taken in upper 20 m during period when the ship stayed in the same water mass (11–19 April).

TABLE 1 Parameters α^B and P_m^B for six cruises in Northwest Atlantic Ocean

Date	Mixed-layer depth (m)	Concentrations in upper 50 m		α^B *	P_m^B †	N	Remarks
		Nitrate (mg-at N m ⁻³)	Chlorophyll (mg m ⁻³)				
April 1983	>80	<0.5	<0.5	0.056 ± 0.006	2.32 ± 0.31	9	Gales. Post-spring-bloom conditions.
August 1985	70	undetectable	~0.25	0.053 ± 0.007	2.35 ± 0.17	19	Summer oligotrophic phase. No DCM‡.
June 1987	~35	<0.2	<0.2	0.053 ± 0.005	1.38 ± 0.16	31	Summer oligotrophic phase. DCM.
September 1988	~50	~0.5	<0.2	0.054 ± 0.007	2.66 ± 0.34	21	Post-autumn bloom? DCM.
April 1989	70	2.5–4.0	1.5–3.0	0.136 ± 0.011	6.17 ± 0.57	21	Spring bloom.
April 1990	~40	1.5 ⇒ 0	~0.8	0.093 ± 0.004	7.46 ± 0.33	51	Declining phase of spring bloom.

Means and standard errors (s.e.) of α^B and P_m^B are given, together with other observations to provide the oceanographic context.

* mg C (mg Chl)⁻¹ h⁻¹ (W m⁻²)⁻¹.

† mg C (mg Chl)⁻¹ h⁻¹.

‡ DCM means deep chlorophyll maximum.

Chl)⁻¹ h⁻¹ and corresponding averages for the deeper samples in a lower range (0.8–2.0). Despite the opposite photoadaptive responses of α^B and P_m^B in deep water, the linear correlation between the two parameters across all six cruises is strong ($r^2=0.72$; $n=152$) (Fig. 3a), indicating a constant value (50 W m⁻²) for the derived parameter $I_k = P_m/\alpha$. This is extremely convenient for the computation of primary production, in that I_k is the commonly used scale factor for normalization of the irradiance in analytical treatments⁸. The correlation between α^B and P_m^B is yet stronger if we consider only the data in the upper layer, say from surface to 20 m (Fig. 3a). In this case, $I_k = 85 \text{ W m}^{-2}$ ($r^2=0.95$; $n=49$), presumably a more representative value for populations not photoadapted to low light.

So far we have discussed only the broad-spectral-band α^B as determined in an incubator excited by a tungsten lamp. For the most detailed calculations of primary production⁹ we have to consider the wavelength-resolved form $\alpha^B(\lambda)$, the photosynthetic action spectrum, which has only recently become accessible to routine measurement at sea¹⁰. We should also consider the wavelength dependence of the optical absorption coefficient for the total particulate material (computation of light penetration¹¹), and for the phytoplankton alone (computation of quantum yield¹²). Within cruises to the same area, we have found that the shapes of these spectra vary little (Fig. 3b): we may use them to estimate the quantum yield of photosynthesis in low light, according to the equation $\phi(B) = \alpha^B(a_*)^{-1}$, where $\langle a_* \rangle_\lambda$ is the specific absorption coefficient for phytoplankton

averaged over the photosynthetically active waveband^{9,13}. In our data, $\langle a_* \rangle_\lambda = 0.019 B^{-0.30}$ ($r^2=0.94$, $n=10$ and B is in mg Chl m⁻³), that is, in the wrong sense to account for the increase in α^B during the bloom. We infer that the seasonal changes in α^B were due to changes in ϕ , perhaps as a consequence of seasonal changes in community composition. Specifically, we find that for spring bloom conditions ($\alpha^B = 0.14$; $B = 2.5$), quantum yield is 45% of its theoretical maximum^{13,14},

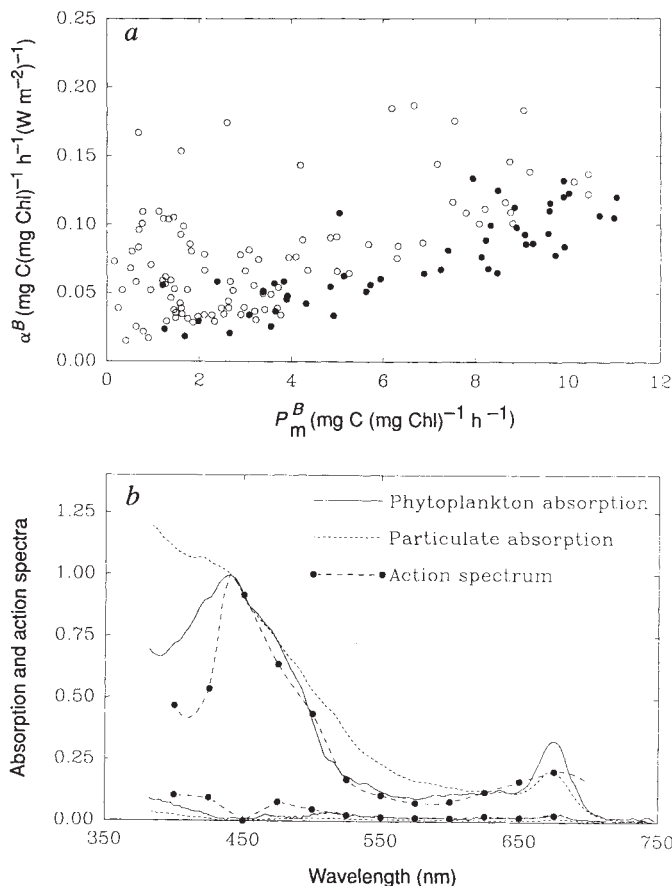


FIG. 3 a, Relationships between α^B and P_m^B for all cruises. Closed circles refer to samples from the upper 20 m of the water column, open circles refer to samples from greater depths. b, Shapes of the action spectrum of photosynthesis (1988 data), and the absorption spectra for phytoplankton and for total particulate material (1989 data). The upper curves are the mean curves for the cruise in question. The lower curves are the standard errors. All curves are normalized to unity at 440 nm, to facilitate comparison of shape rather than magnitude. The absorption spectra were measured on filters according to the method of ref. 15, as modified in refs 16 and 17. They were calibrated by comparison of simultaneous measurements of absorption on filter and in suspension using laboratory cultures.

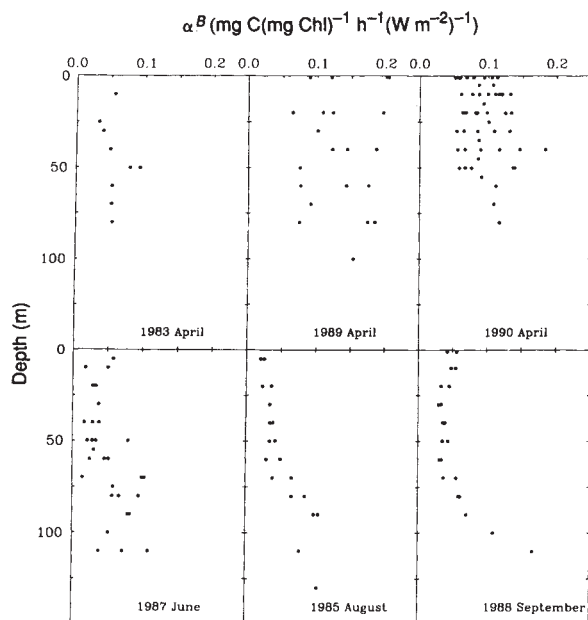


FIG. 2 Depth dependence of α^B for the six cruises, showing the difference between spring (unstratified) conditions and summer (stratified) conditions.

whereas in post-bloom conditions ($\alpha^B = 0.05$; $B = 0.2$), it is only 7% of the theoretical maximum. Such changes¹² in ϕ should certainly be taken into account in modelling the onset of the spring bloom in the ocean for those models based on absorbed rather than available light⁹.

These findings underline the importance of blooms in the carbon economy of the ocean. Using equations (26) and (27) of ref. 8, we can calculate the sensitivity of the daily water-column production to variations in α^B and P^B respectively. Taking a value of $I_k = 85 \text{ W m}^{-2}$ and a nominal, noon surface irradiance of $I_0^m = 500 \text{ W m}^{-2}$, we find a scaled, noon irradiance $I^m = I_0^m / I_k \sim 6$. Then, a 100% change in α^B , such as observed between the spring bloom phase and the oligotrophic summer phase, would lead to a 50% change in estimated daily production, definitely worth correcting for. On the other hand, the variation in α^B within a season is small enough that the assumption of constant α^B would lead to negligible bias in the estimate of daily production. Similar remarks apply to P_m^B . As we have emphasized elsewhere⁴, the fluctuations in photosynthesis parameters within provinces and seasons are less important for calculation of primary production than those in the photo-

synthetic biomass, which has a much wider dynamic range and for which data can be collected by satellite.

Many of the fundamental questions before oceanographers today, particularly those relating to the planetary carbon cycle, require answers framed at the global or basin scale. At these scales, synoptic coverage by ship is out of the question, and one is obliged to resort to the synoptic (but imperfect) information provided by satellites, supplemented by oceanographic insights obtainable only by ships. Such an approach will not work unless it is possible to partition the oceans into a restricted set of 'biogeochemical provinces' within which the key rate parameters of the pelagic ecosystem (including the photosynthesis parameters) can be considered constant in a given season^{1,3,4}. Our results encourage this belief and they establish the North Sargasso Sea as belonging to one such province whose boundaries remain to be defined. Following a partition-by-provinces approach does not imply a lack of interest in understanding how environmental fluctuations affect ecological fluxes. Both lines should be pursued: we expect that they will converge more readily if the focus is on the parameters that control the key fluxes, rather than on the fluxes themselves. □

Received 31 July 1991; accepted 11 February 1992.

- Platt, T. & Sathyendranath, S. *Science* **241**, 1613-1620 (1988).
- Fasham, M. J. R., Ducklow, H. W. & McKelvie, S. M. *J. mar. Res.* **48**, 591-639 (1990).
- Mueller, J. L. & Lange, R. E. *Limnol. Oceanogr.* **34**, 1572-1586 (1989).
- Platt, T., Caverhill, C. & Sathyendranath, S. *J. geophys. Res.* **96**, 15, 147-15, 159 (1991).
- Platt, T., Denman, K. L. & Jassby, A. D. in *The Sea. Ideas and Observations on Progress in the Study of the Seas*, Vol. VI (ed. Goldberg, E. D.) 807-856 (Wiley, New York, 1977).
- Platt, T., Gallegos, C. L. & Harrison, W. G. *J. mar. Res.* **38**, 687-701 (1980).
- Lewis, M. R., Horne, E. P. W., Cullen, J. J., Oakey, N. S. & Platt, T. *Nature* **311**, 49-50 (1984).
- Platt, T., Sathyendranath, S. & Ravindran, P. *Proc. R. Soc. B241*, 101-111 (1990).
- Sathyendranath, S. & Platt, T. *Limnol. Oceanogr.* **34**, 188-198 (1989).
- Lewis, M. R., Warnock, R., Irwin, B. & Platt, T. *J. Phycol.* **21**, 310-315 (1985).

- Sathyendranath, S. & Platt, T. *J. geophys. Res.* **93**, 9270-9280 (1988).
- Cleveland, J. S., Perry, M. J., Kiefer, D. A. & Talbot, M. C. *J. mar. Res.* **47**, 869-886 (1989).
- Platt, T. & Jassby, A. D. *J. Phycol.* **12**, 421-430 (1976).
- Kirk, J. T. O. *Light and Photosynthesis in Aquatic Ecosystems* (Cambridge University Press, UK, 1983).
- Yentsch, C. S. *Limnol. Oceanogr.* **7**, 207-217 (1962).
- Mitchell, B. G. & Kiefer, D. A. *Deep Sea Res.* **35**, 665-689 (1988).
- Kishino, M., Okami, N., Takahashi, M. & Ichimura, S. *Limnol. Oceanogr.* **31**, 557-566 (1986).

ACKNOWLEDGEMENTS. This work was supported by the Office of Naval Research (USA), NASA (USA) and the Department of Fisheries & Oceans, Canada. Additional support was provided by the Natural Sciences and Engineering Research Council through Operating Grants to S.S. and T.P.

A linear free energy relationship for crystalline solids and aqueous ions

Dimitri A. Sverjensky* & P. A. Molling†

* Department of Earth and Planetary Sciences, The Johns Hopkins University, Baltimore, Maryland 21218, USA

† Unocal Geothermal Division, Unocal Corporation, PO 6854, Santa Rosa, California 95406, USA

QUANTITATIVE chemical modelling of geochemical, environmental and industrial processes is often severely limited because of large gaps in experimentally derived thermodynamic databases for Gibbs free energies of crystalline solids and aqueous ions. Methods proposed previously for estimation of the free energies or enthalpies of formation of crystalline solids¹⁻⁹ are subject to large uncertainties, typically greater than ± 5 -10 kcal mol⁻¹. Here we present an empirically based linear free energy equation¹⁰ applicable to cations of any charge, radius or chemical type, which allows estimates of the free energies of solids with uncertainties of less than ± 1 kcal mol⁻¹. Our equation is analogous to the linear free energy relations of Hammett and others^{11,12} for aqueous organic reactions, but applies instead to crystalline solids. We apply our equation to experimentally derived standard Gibbs free energies of formation of isostructural families of divalent oxides, hydroxides, carbonates, fluorides, chlorides and sulphates. This new level of accuracy for predicting free energies of crystalline solids opens up opportunities for chemical modelling of many processes not previously amenable to thermodynamic analysis.

Chemical modelling of processes such as the partitioning of trace elements between solids, or between solids and aqueous fluids (coprecipitation), requires standard Gibbs free energies of formation that are often difficult or impossible to measure

experimentally (examples are SrCO₃ with the calcite structure and MnAl₂Si₂O₈ with the anorthite structure). Uncertainties in estimating the necessary free energies or enthalpies¹⁻⁹ propagate through calculated standard free energies of reaction and lead to enormous uncertainties in predicted equilibrium constants. An alternative approach involves correlating the standard Gibbs free energies of formation of an isostructural family of crystalline solids with those of aqueous cations of a given charge¹³. This eliminates much of the scatter associated with estimation techniques proposed previously. Within an isostructural family of solids, the free energy change in going from one compositional endmember to another containing cations with the same (or very similar) ionic radii is paralleled by the free energy change between the corresponding aqueous cations. Such a relationship is directly analogous to the Hammett linear free energy relationship for aqueous organic reactions¹¹⁻¹², but here the substituents are the cations (D.A.S., manuscript in preparation). Also, because we are dealing with crystalline solids, the effects of crystallographic radius must be accounted for explicitly. We apply these concepts to crystalline solids containing divalent cations as follows.

We first represent the compositional endmembers of an isostructural family of solids by the chemical formula M_nX, where M²⁺ is a divalent cation and X represents the remainder of the composition of the solid (X is in general polyatomic; for example, metal carbonates with the calcite structure, where M is Mg, Ca or Fe and X is CO₃). With experimentally derived standard Gibbs free energies of formation of the endmember solids ($\Delta G_{f,M,X}^0$), a linear free energy correlation is given by the equation

$$\Delta G_{f,M,X}^0 = a_{M,X} \Delta G_{n,M^{2+}}^0 + b_{M,X} + \beta_{M,X} r_{M^{2+}} \quad (1)$$

In this equation, $a_{M,X}$, $b_{M,X}$ and $\beta_{M,X}$ are coefficients characteristic of the particular crystal structure represented by M_nX, and $r_{M^{2+}}$ represents the Shannon-Prewitt radius¹⁴ (in ångströms) of M in a given coordination state. The parameter $\Delta G_{n,M^{2+}}^0$

TABLE 1 Ionic radii and thermodynamic data for aqueous cations and eight families of solids

M	$r_{M^{2+}}$ (Å)*	ΔG_f										
		ΔG_s $M^{2+}(\text{aq})$	$\Delta G_f^\dagger$ $M^{2+}(\text{aq})$	ΔG_n $M^{2+}(\text{aq})$	MO‡ (NaCl)	M(OH) ₂ § (CdCl ₂)	MF ₂ (TiO ₂)	MF ₂ ¶ (CaF ₂)	MCl ₂ # (CdCl ₂)	MCO ₃ ** (calcite)	MCO ₃ †† (arag.)	MSO ₄ ‡‡ (BaSO ₄)
Be	0.45	-175.02	-89.80¶	85.23								
Mg	0.72	-145.80	-108.83	36.97	-136.04	-199.54	-255.78		-141.44	-246.1		
Ca	1	-121.28	-132.12	-10.83	-144.2	-214.51		-280.97		-269.88	-269.68	
Mn	0.82	-136.46	-55.20	81.26	-86.74	-147.86	-192.3		-105.26	-195.56		
Fe	0.77	-141.04	-21.87	119.17		-117.73				-162.41		
Co	0.735	-144.35	-13.00	131.35	-51.2	-109.61	-148.36		-64.48			
Ni	0.7	-147.75	-10.90	136.85	-50.6	-109.56	-145.87		-61.91			
Cu	0.73	-144.83	15.55	160.38								
Zn	0.745	-143.40	-35.17	108.23			-170.48			-174.84		
Sr	1.16	-109.30	-133.72	-24.41	-137.12			-278.39			-272.6	-320.7
Cd	0.95	-125.31	-18.57	106.74		-113.34		-154.8	-82.2	-159.99		
Sn	1.11	-112.91	-6.63	106.28								
Ba	1.36	-95.99	-132.73	-36.73	-124.37			-271.57			-270.6	-324.26
Eu	1.17	-108.59	-129.10	-20.50	-133.1							
Hg	1.02	-119.71	939.36	159.07								
Pb	1.18	-107.89	-5.79	102.10							-149.46	-194.51
Ra	1.39	-94.14	-134.20	-40.06								-326.29
UO ₂	0.754	-142.54	-227.70	-85.15								

Experimental ionic radii (sixfold coordination) for divalent cations in crystalline solids ($r_{M^{2+}}$) and standard Gibbs free energies of formation (ΔG_f) of divalent aqueous cations $M^{2+}_{(\text{aq})}$ and crystalline solids M_nX , together with calculated Born solvation free energies (ΔG_s) and values of ΔG_n . All values refer to 25 °C and 1 bar and are in kcal mol⁻¹ unless otherwise stated. Values of $r_{M^{2+}}$ are from ref. 14 unless otherwise stated. ΔG_s of $M^{2+}_{(\text{aq})}$ are calculated from the given radii with equations (3)–(6). $\Delta G_f M^{2+}_{(\text{aq})}$ from CODATA²² unless otherwise stated. ΔG_n from equation (2). Structure types of solids are given in parentheses.

* Be from ref. 23; Sn from D. Sassani and E. L. Shock (personal communication); Ra estimated from eightfold and twelvefold radii in ref. 23 and radii for Ba²⁺; UO₂ calculated from ref. 15 and entropy S from ref. 22.

† Be, Mn from ref. 17; Fe from ref. 24; Co, Ni, Sr, Eu and Ra from ref. 16; Ba from ref. 26.

‡ Mg, Ca calculated from CODATA²²; Mn, Co, Ni, Eu from ref. 16; Sr from refs 17, 25; Ba from ref. 26.

§ Mg obtained from analysis of high-temperature phase equilibria²⁷ consistent with MgO in ref. 22; Ca from ref. 28; Mn, Co, Ni, Cd from stable, aged or inactive value of log $K_{(\text{solubility})}$ in ref. 29. Fe from active value of log $K_{(\text{solubility})}$ in ref. 29.

|| Mg, Zn from ref. 16; Mn, Co, Ni from ref. 17.

¶ Ca from ref. 28; Sr, Cd from ref. 16; Ba from ref. 17.

Reference 16.

** Mg from analysis of high-temperature phase equilibria²⁷, consistent with MgO in ref. 22; Ca from ref. 30; Mn from ref. 31; Fe calculated from solubility product in ref. 32 using revised free energy of the aqueous ferrous ion tabulated here; Zn, Cd from ref. 16.

†† Ca from ref. 30; Sr calculated from solubility product of strontianite (log $K = -9.271$)³³; Ba from ref. 26; Pb from ref. 34.

‡‡ Sr, Ba calculated from solubility product in ref. 35; Ra from solubility product in ref. 36; Pb from ref. 17.

represents a radius-based correction to the standard Gibbs free energy of formation of the aqueous cation M^{2+} ($\Delta G_{f,M^{2+}}^0$) calculated using

$$\Delta G_{n,M^{2+}}^0 = \Delta G_{f,M^{2+}}^0 - \Delta G_{s,M^{2+}}^0 \quad (2)$$

where $\Delta G_{s,M^{2+}}^0$ denotes the standard Gibbs free energy of solvation of the aqueous cation¹⁵. We calculated solvation free energies from conventional Born solvation coefficients for the aqueous cations ($\omega_{M^{2+}}$) according to the equation¹⁵

$$\Delta G_{s,M^{2+}}^0 = \omega_{M^{2+}}(1/\epsilon - 1) \quad (3)$$

Where ϵ is the dielectric constant of water (78.47 at 25 °C and 1 bar). In turn, the conventional Born solvation coefficients were calculated from the equation¹⁵

$$\omega_{M^{2+}} = \omega_{M^{2+}}^{\text{abs}} - 2\omega_{H^+}^{\text{abs}} \quad (4)$$

where $\omega_{H^+}^{\text{abs}}$ is 0.5387×10^5 cal mol⁻¹ (ref. 15). Values of the absolute Born coefficient ($\omega_{M^{2+}}^{\text{abs}}$) were obtained from effective electrostatic radii of the aqueous cations ($r_{e,M^{2+}}$) used in the equations¹⁵

$$\omega_{M^{2+}}^{\text{abs}} = (1.66027 \times 10^5)4/(r_{e,M^{2+}}) \quad (5)$$

and

$$r_{e,M^{2+}} = r_{M^{2+}} + 2(0.94) \quad (6)$$

where $r_{M^{2+}}$ represents the crystallographic radius of the cation M^{2+} . We found that Shannon–Prewitt radii¹⁴ referring to sixfold coordination (Table 1) gave optimal results in the regression calculations described below. The above equations were used to calculate values of $\Delta G_{n,M^{2+}}^0$ (Table 1) for eighteen divalent aqueous cations ($Z = 2$).

Equation (1) closely fits the experimentally derived standard Gibbs free energies of formation of many isostructural families

of solids, including oxides, hydroxides, carbonates, fluorides, chlorides, iodides, sulphates, sulphides, phosphates and silicates. Shown in Table 1 are experimental values of the Gibbs free energies of formation of members of eight isostructural families of solids representing available data for simple oxides, hydroxides, carbonates, fluorides, chlorides and sulphates. These families (for which we have the most data) provide an excellent illustration of the linear free energy method. Other isostructural families that have been investigated include the bromides, iodides, sulphides, phosphates, perovskite oxides, and olivine and garnet-type silicates (D.A.S., manuscript in preparation). The experimental data in Table 1 are primarily from standard compilations. Considerable care has been taken to ensure internal consistency of the thermodynamic data. The free energies of the solids belonging to each isostructural family were regressed using equation (1) together with values of ionic radii and $\Delta G_{n,M^{2+}}^0$ from Table 1. The regression results for the eight isostructural families are given in Table 2 (uncertainties are given in parentheses for each regression parameter). Calculated free energies are given in Table 3, and the differences between experimental and calculated values of the free energies of formation of the solids are shown in Fig. 1. For 40 out of the 43 points plotted, the discrepancy between calculated and observed free energies is less than 1.0 kcal mol⁻¹. The worst discrepancies are for MnO, Cd(OH)₂ and ZnF₂, all of which are within ± 1.5 kcal mol⁻¹. Of all the thermodynamic data examined for the eight isostructural families in Fig. 1, only the free energy data for CdO were excluded from the regression analysis because of an anomalously large discrepancy between calculated and measured data. In each of the other three cases, examination of the original experimental sources and thermodynamic cycles reveals significant uncertainties in the free energy data for the aqueous species (for example Mn²⁺; see refs 16 and 17, which differ by 1 kcal mol⁻¹) and/or free energy data

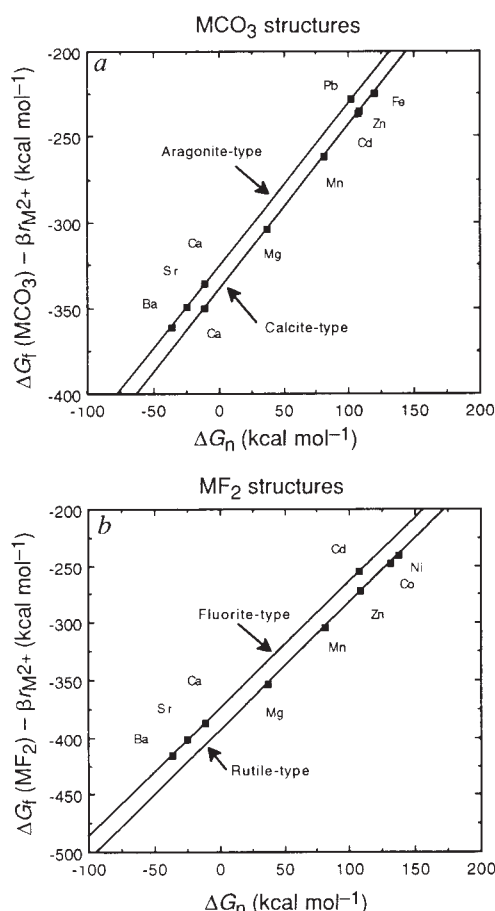


FIG. 2 Graphical representation of equation (7) for pairs of polymorphic structures. The symbols represent experimentally derived values of the standard Gibbs free energies of formation from Table 1, from which we subtracted the term $\beta_{M,X} r_{M^{2+}}$, using the appropriate values of $r_{M^{2+}}$ and $\beta_{M,X}$ from Tables 1 and 2, respectively. Values of the cation parameter $\Delta G_{nM^{2+}}^0$ were taken from Table 1. The lines in each graph represent calculated values generated with equation (7) and regression coefficients from Table 2. Because the lines for the two carbonate structures are almost parallel, and the lines for the two fluoride structures are also almost parallel, the regression results suggest that the parameter $a_{M,X}$ (see Table 2) is characteristic not just of a single crystal structure but of all polymorphs of the composition M_X . This should permit accurate estimation of this parameter in situations where there are limited data available. Note, however, that the regression parameter $\beta_{M,X}$ is different for different crystal structures (see Table 2).

parameters $a_{M,X}$, $b_{M,X}$ and $\beta_{M,X}$, and systems involving more complex oxides such as the spinel and the perovskite types, and silicate minerals. Correlations such as those discussed above can be used to predict the free energies of numerous transition metal-bearing spinels, perovskites, silicates and other crystalline solids for which experimental data are currently unavailable. □

- Shock, E. L. & Helgeson, H. C. *Geochim. cosmochim. Acta* **52**, 2009–2036 (1988).
- Wagman, D. D. et al. *J. Phys. Chem. Ref. Data* Vol. 11, Suppl. 2 (1982).
- Glushko, V. P. *Thermal Constants of Substances* (VINITI, Moscow, 1965–1981).
- Veizer, J. *Reviews in Mineralogy*, Vol. 11, 265–300 (Mineralogical Society of America, Chelsea, Michigan, 1983).
- Mucci, A. & Morse, J. W. *Aquat. Sci.* **3**, 217–254 (1990).
- Sverjensky, D. A. *Geochim. cosmochim. Acta* **48**, 1127–1134 (1984).
- Mann, A. W. & Deutscher, R. L. *Econ. Geol.* **73**, 1724–1737 (1978).
- Cox, J. D., Wagman, D. D. & Medvedev, V. A. *CODATA Key Values for Thermodynamics* (Hemisphere, New York, 1989).
- Shannon, R. D. *Acta Cryst.* **A32**, 751–767 (1976).
- Shock, E. L. & Helgeson, H. C. *Geochim. cosmochim. Acta* **53**, 215–216 (1989).
- Mah, A. D. *US Bureau of Mines Rep. Invest.* 6171 (1963).
- Busenberg, E. & Plummer, L. N. *Geochim. cosmochim. Acta* **50**, 2225–2234 (1986).
- Berman, R. G. *J. Petrol.* **29**, 445–522 (1988).
- Garvin, D., Parker, V. B. & White, H. L. Jr. *CODATA Thermodynamic Tables—Selections for Some Compounds of Calcium and Related Mixtures: A Prototype Set of Tables* (Hemisphere, New York, 1987).
- Baes, C. F. & Mesmer, R. E. *The Hydrolysis of Cations* (Wiley, New York, 1976).
- Plummer, L. N. & Busenberg, E. *Geochim. cosmochim. Acta* **46**, 1011–1040 (1982).
- Robie, R. A., Haselton, H. T. Jr & Hemingway, B. S. *Am. Miner.* **69**, 349–357 (1984).
- Langmuir, D. *US Geol. Survey Prof. Pap.* 650-B, B180–184 (1969).
- Busenberg, E., Plummer, L. N. & Parker, V. B. *Geochim. cosmochim. Acta* **48**, 2021–2036 (1984).
- Robie, R. A., Hemingway, B. S. & Fisher, J. R. *US Geol. Survey Bull. no.* 1452 (1978).
- Nordstrom, D. K. et al. *Am. chem. Soc. Symp. Ser.*, Ch. 31, 398–413 (1990).
- Langmuir, D. & Riese, A. C. *Geochim. cosmochim. Acta* **49**, 1593–1602 (1985).

ACKNOWLEDGEMENTS. We thank H. C. Helgeson, W. F. McKenzie, A. Komninou, E. Shock and D. Veblen for discussions, and N. Scrivner for support and encouragement. Financial support was provided by the NSF, the American Chemical Society (Petroleum Research Fund), and Du Pont.

Effect of animal husbandry on herbivore-carrying capacity at a regional scale

M. Oesterheld*, O. E. Sala* & S. J. McNaughton†

* Departamento de Ecología, Facultad de Agronomía, Universidad de Buenos Aires, Av. San Martín 4453, 1417 Buenos Aires, Argentina

† Biological Research Laboratories, Syracuse University, 130 College Place, Syracuse, New York 13244-1220, USA

ALL significant properties of the herbivore trophic level, including biomass, consumption and productivity, are significantly correlated with primary productivity across a broad range of terrestrial ecosystems^{1,2}. Here we show that livestock biomass in South American agricultural ecosystems across a 25-fold gradient of primary productivity exhibited a relationship with a slope essentially identical to unmanaged ecosystems, but with a substantially greater y-intercept. Therefore the biomass of herbivores supported per unit of primary productivity is about an order of magnitude greater in agricultural than in natural ecosystems, for a given level of primary production. We also present evidence of an increase in livestock body size with primary productivity, a pattern previously characterized in natural ecosystems³. To our knowledge this is the first quantitative documentation at a regional scale of the impact of animal husbandry practices, such as herding, stock selection and veterinary care, on the biomass and size-structure of livestock herds compared with native herbivores.

Our objectives were to assess the impact of human intervention on herbivore carrying capacity, to determine whether the close relationship between natural herbivore biomass and primary productivity observed across terrestrial habitats^{1,2} reflects a limitation imposed by primary production, and to assess the impact of human intervention on the relationship between herbivore body size and primary productivity. We searched for a geographic region where (1) rangelands grazed by domestic herbivores are dominant, (2) primary production varies over a broad range, and (3) management practices are limited to elementary animal husbandry. The southern portion of South America meets those criteria. Together with the North American Great Plains, Asian grasslands and African savannas, it represents one of Earth's largest expanses of natural rangelands. South American rangelands encompass several vegetation types: desert, steppe, subhumid temperate grassland and sub-

Received 21 October 1991; accepted 11 February 1992.

- Tardy, Y. & Garrels, R. M. *Geochim. cosmochim. Acta* **40**, 1051–1056 (1976).
- Tardy, Y. & Garrels, R. M. *Geochim. cosmochim. Acta* **41**, 87–92 (1977).
- Helgeson, H. C., Delaney, J. M., Nesbitt, H. W. & Bird, D. C. *Am. J. Sci.* **A278**, 1–229 (1978).
- Viellard, P. H. & Tardy, Y. *Am. J. Sci.* **288**, 997–1040 (1988).
- Hazen, R. M. *Am. J. Sci.* **288A**, 242–269 (1988).
- Essington, M. E. *Soil. Sci. Soc. Am. J.* **52**, 1574–1579 (1988).
- Cherniak, J. A. & Rimstidt, J. D. *Am. Miner.* **74**, 1023–1031 (1989).
- Cherniak, J. A. & Rimstidt, J. D. *Am. Miner.* **75**, 1376–1380 (1990).
- Zugev, V. V. *Geochim. Int.* **2**, 45–51 (1989).
- Sverjensky, D. A. *Geol. Soc. Am. Ann. Meeting Abstr.* **23**, A212 (1991).
- Wells, P. R. *Linear Free Energy Relationships* (Academic, London, 1968).
- Exner, O. *Correlation Analysis of Chemical Data* (Plenum, New York, 1988).
- Sverjensky, D. A. *Geochim. cosmochim. Acta* **64**, 853–864 (1985).
- Shannon, R. D. & Prewitt, C. T. *Acta Crystallogr.* **B25**, 925–946 (1969).

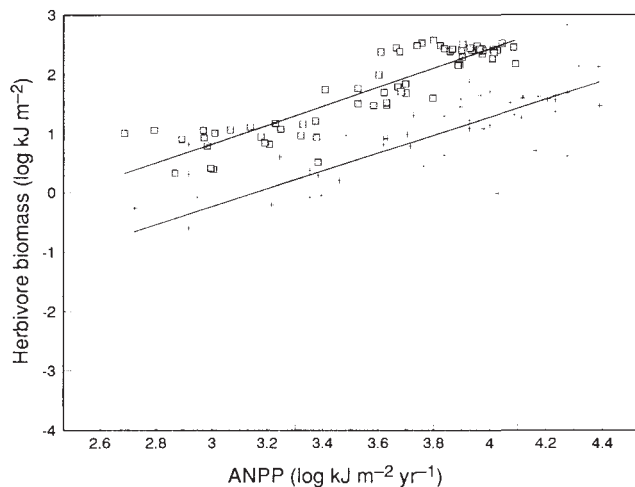


FIG. 1 Relationship between net above-ground primary productivity (ANPP) and herbivore biomass for natural (crosses) and agricultural (squares) ecosystems. The data set consisted of 67 agricultural points and 51 natural points. The latter were taken from refs 1 and 2. Logarithms are on a decimal base. Energy-biomass conversion units are 9,900 kJ per kg livestock fresh weight²⁰ and 16.76 kJ per gram of ANPP (ref. 21).

tropical savanna. They are used principally by domestic herbivores, cattle (*Bos indicus*) and sheep (*Ovis aries*). The technological level of husbandry practices is more developed than pastoralism, but does not reach the intensities of care characteristic of Europe or North America. Animals are rarely fed silage or grain, and management practices that tend to increase primary productivity, such as fertilization, irrigation and forage species replacement, are rare. But effort is devoted to the improvement of animal status: herds are organized by categories, ranches are subdivided in paddocks, water holes are abundant, veterinary care is intense, predators and parasites are controlled, and breed purity is highly appreciated.

We compiled a data set of livestock biomass from censuses at the county level across Argentina and Uruguay^{4,5}. Above-ground net primary productivity (ANPP) was estimated from annual rainfall data⁶ using a linear model⁷ that closely fits primary productivity data for 14 sites in South America⁸. Our data set was compared with recently published data on herbivore biomass, from insects to large ungulates, in natural terrestrial habitats^{1,2}.

Livestock biomass, B (kJ m^{-2}), was associated with ANPP ($\text{kJ m}^{-2} \text{yr}^{-1}$; $r^2 = 0.797$, $P < 0.000001$, d.f. = 65) by $\log B = 1.602 \log \text{ANPP} - 3.98$. A comparison of this pattern with the best-fit line from natural ecosystems (Fig. 1) reveals similarity of slopes, but a significant difference in y -intercept. The slope for natural habitats was 1.52, not significantly different from the slope for agricultural ecosystems ($P = 0.706$, d.f. = 1, 114). The y -intercept for natural habitats was -4.79 , significantly lower than that for livestock ($P < 0.00001$, d.f. = 1, 115). Livestock biomass supported per unit of primary production was about an order of magnitude greater than herbivore biomass in natural ecosystems.

In natural ecosystems, increasing herbivore biomass with primary productivity is accompanied by an increase in average herbivore body size³. A similar pattern was documented for livestock in our data set (Fig. 2). The proportion of sheep biomass decreased, whereas the proportion of cattle increased along the productivity gradient. This is not the result of compulsory management techniques. Therefore it must be the result of ecological constraints, but we are not certain about the ecological determinants of this pattern.

The results have four major implications for ecosystem theory and agricultural practice. First, the exponential relationship between livestock biomass and primary productivity described here provides a quantitative reference for those attempting to determine the long-term carrying capacity of a system, a key step in rangeland management⁹.

Second, the data suggest that although herbivore biomass and primary productivity are closely related in natural ecosystems, primary productivity is not the only limiting factor (see ref. 10

and refs therein). Determining the factors influencing the carrying capacity of ecosystems has been a major concern of ecology since its beginnings^{11,12}, and is an obvious pursuit of range management. Elementary management techniques aimed at providing drinking water and minerals, regulating animal distribution and reducing parasitism, predation and disease have a large impact on ecosystem-carrying capacity. In agreement with our results, it has been shown that herbivore biomass and rainfall in African systems are related¹³, and that agricultural and pastoral systems have greater herbivore biomass than their natural counterparts^{13,14}.

Third, the results can contribute to resolving a contradiction between suggestions that a high density of wild ungulates may improve forage resources¹⁵ but a high density of livestock depletes them¹⁶. It has been proposed that grazing by herds of wild ungulates results in the formation of grazing lawns that favour forage intake due to higher forage concentration per unit volume, higher nutritional quality and reduced mastication time¹⁵. This claim is contradicted by evidence from agroecosystems showing that high herbivore density leads to a deterioration of forage resources¹⁶. Our results indicate that 'high density' has a very different meaning when comparing the two types of ecosystem as there is a tenfold difference in herbivore load between them.

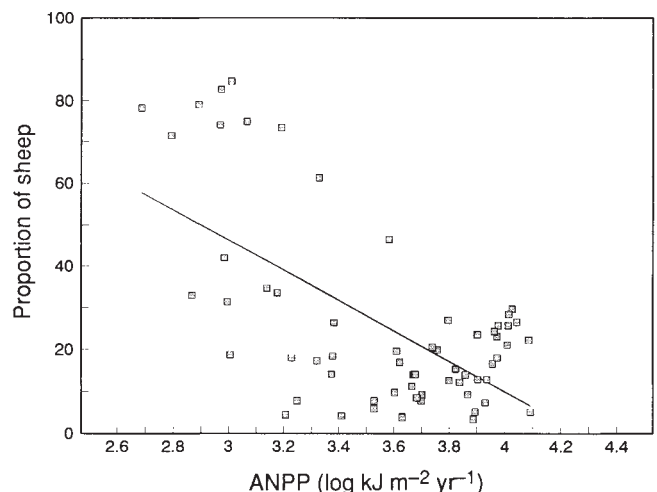


FIG. 2 Relationship between the angular-transformed proportion of sheep biomass and ANPP ($\text{kJ m}^{-2} \text{yr}^{-1}$) across a variety of rangeland agroecosystems in southern South America. Proportions were angularly transformed to homogenize variances. The best-fit line was $y = 155.9 - 36.51 \log \text{ANPP}$ ($r^2 = 0.38$, $P < 0.000001$, d.f. = 65).

Finally, the results can explain the greater grazing effects on grassland species composition in South America compared with those in Africa. Subhumid grasslands of South America are invaded by exotic, mostly European, species when grazed by livestock and the importance of native grasses is drastically reduced¹⁷. In contrast, equivalent subhumid grasslands in East African game reserves are always dominated by native grasses, even under high grazing pressure¹⁸. A general model of the impact of grazing on community structure¹⁹ attributed this difference to the longer evolutionary history of grazing in African grasslands and savannas compared with those in South America, whose grazer and browser fauna almost totally failed to survive the Pleistocene extinctions of large herbivorous mammals. Our results show that grazing evolutionary history is not the only major difference between these systems. The much higher herbivore load supported by South American agricultural grasslands over the past two centuries may contribute to their different response to grazing. □

Received 7 October; accepted 31 December 1991.

- McNaughton, S. J., Oesterheld, M., Frank, D. A. & Williams, K. J. *Nature* **341**, 142–144 (1989).
- McNaughton, S. J., Oesterheld, M., Frank, D. A. & Williams, K. J. in *Comparative Analyses of Ecosystems: Patterns, Mechanisms and Theories* (eds Cole, J. J., Lovett, G. M. & Findlay, S. E. G.) 120–139 (Springer, New York, 1991).
- East, R. *Afr. J. Ecol.* **22**, 245–270 (1984).
- Ministerio de Economía Empadronamiento Nacional Agropecuario y Censo Ganadero (Buenos Aires, Argentina, 1974).
- Ministerio de Agricultura y Pesca *Censo General Agropecuario* (Montevideo, Uruguay, 1980).
- Production Yearbook 1986* (United Nations Food and Agriculture Organization, Rome, 1987).
- Lauenroth, W. K. in *Perspectives in Grassland Ecology* (ed. French, N. R.) 3–24 (Springer, New York, 1979).
- McNaughton, S. J., Sala, O. E. & Oesterheld, M. in *Biological Relationships Between Africa and South America* (ed. Goldblatt, P.) (Yale Univ. Press, New Haven, in the press).
- Scarnecchia, D. L. *J. Range Manage.* **43**, 553–555 (1990).
- Moen, J. & Oksanen, L. *Nature* **353**, 510 (1991).
- Darwin, C. *The Origin of Species* 6th London edn (Thompson & Thomas, Chicago, 1872).
- Errington, P. L. *Science* **124**, 304–307 (1956).
- Coe, M. J., Cumming, D. H. & Phillipson, J. *Oecologia* **22**, 341–354 (1976).
- Sharkey, M. J. *Mammalia* **34**, 564–572 (1970).
- McNaughton, S. J. *Am. Nat.* **124**, 863–886 (1984).
- Westoby, M. *Am. Nat.* **126**, 870–871 (1985).
- Sala, O. E., Oesterheld, M., León, R. J. C. & Soriano, A. *Vegetatio* **67**, 27–32 (1986).
- McNaughton, S. J. *Ecol. Monogr.* **53**, 291–320 (1983).
- Milchunas, D. G., Sala, O. E. & Lauenroth, W. K. *Am. Nat.* **132**, 87–106 (1988).
- Coughenour, M. B. *et al. Science* **230**, 619–625 (1985).
- Golley, F. B. *Ecology* **42**, 581–584 (1961).

ACKNOWLEDGEMENTS. We thank C. Di Bella for help with gathering data. Supported by the NSF, CONICET, and Fundación Antorchas. This research was stimulated by an invitation to participate in a Missouri Botanical Garden Symposium on 'Biological Relationships Between Africa and South America.'

Time to collision is signalled by neurons in the nucleus rotundus of pigeons

Yongchang Wang & Barrie J. Frost*

Departments of Physiology and Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada

THROUGHOUT the animal kingdom, the sight of a rapidly approaching object usually signals danger and elicits an escape response^{1–6}. Gibson⁷ suggested that the symmetrical expansion of an object's image (looming) is the critical variable determining that the object is on a collision course with the observer. Similarly, large expanding flow-fields like those produced by locomotion may precipitate manoeuvres such as turning or landing^{8,9}. From such observations it has been shown that the optic flow parameter, τ , which specifies time to contact with the approaching object best fits the behavioural data^{10,11}. We describe a subpopulation of neurons in the nucleus rotundus of the pigeon brain that respond selectively to objects moving on a collision course towards the bird.

These neurons give their maximum response at a constant time before contact occurs, even when the size of the stimulus or its velocity is varied widely. We propose that these neurons are signalling the time to collision of approaching objects.

We examined the response properties of neurons in the pigeon nucleus rotundus, a major midbrain nucleus in the tectofugal visual pathway, to various computer-generated visual displays. Here we report the responses of neurons in the dorsal posterior zone of the nucleus rotundus, where preliminary qualitative observations had indicated some neurons might respond to motion in depth¹². A stimulus pattern resembling a soccer ball, consisting of a sphere with many alternating black and white panels, was generated by a graphics computer and projected onto a tangent screen 40 cm in front of the bird by a high-

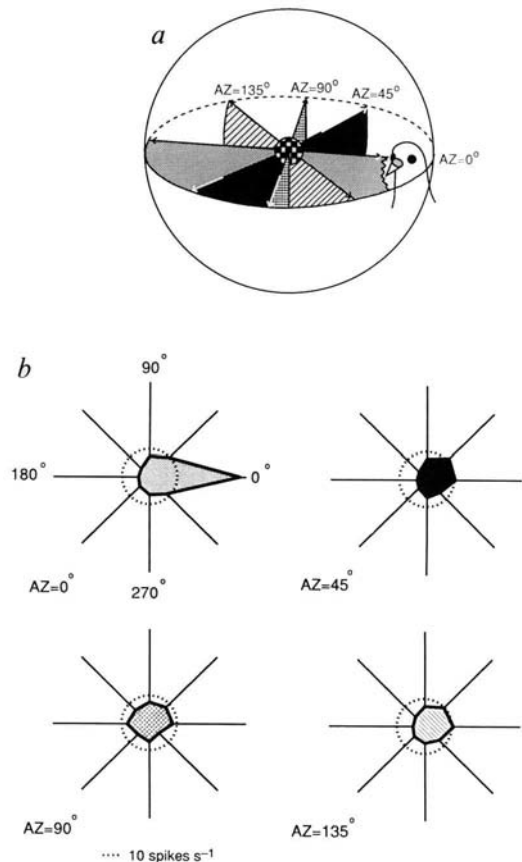


FIG. 1 A single neuron from the dorsal posterior zone of the nucleus rotundus exhibits clear selectivity for a looming visual stimulus. *a*, Diagram to illustrate the 4 planes along which a soccer-ball shaped stimulus pattern, consisting of black and white panels presented monocularly, was moved along 26 simulated three-dimensional trajectories 45° apart in spherical coordinates. (The 4 planes each with 8 directions total 32, but the vertical up and down directions are common, thus resulting in 26 directions.) *b*, Firing rate plots for the different directions of motion of the spherical stimulus within these planes. Each direction of motion was presented 5 times in a randomly interleaved sequence to a pigeon, anaesthetized with ketamine/xylazine (50 mg per kg, 5 mg per kg, respectively), and the values plotted represent the mean firing rate for each three-dimensional direction. An IRIS 4d/310 GTX graphics computer was used to generate the stimulus which was projected by a 1280 × 1024 pixel RGB projector. In the standard X-Y plot (tangent screen plane, or azimuth (AZ)=90°), there is no indication of directional preference, and firing rate is low. For the 0° azimuthal plane (z-axis) there is a strong preference for stimuli directly approaching the bird on a collision course. Polar tuning plots for directions specifying the azimuthal 45° and 135° planes similarly show no strong directional preference. This pattern of activity was typical of the 24 neurons studied. If the origin of the spherical coordinate system was moved to a different position in the receptive field, the direct collision direction again produced the maximal response in the looming-sensitive cells.

* To whom correspondence should be addressed at Department of Psychology.

resolution projector. This system allows realistic depiction of three-dimensional space and simulation of the motion of objects through this space or, alternatively, motion of the observer through this space.

Cells from the dorsal posterior zone (histologically confirmed) of the nucleus rotundus of 20 anaesthetized pigeons were studied quantitatively using standard extracellular recording techniques^{13,14}. Data were recorded for each cell during motion of the spherical stimulus pattern along 26 three-dimensional trajectories spaced 45° apart in spherical coordinates (Fig. 1a). Their monocular receptive fields, determined by hand-moved stimuli, were extremely large (approximately 110° of arc). Of 145 cells, 24 (16.5%) gave maximum responses to a looming stimulus moving towards the animal on a collision course (0° azimuth, 0° elevation), and very little response to motion in other directions through three-dimensional space, as illustrated by a typical cell in Fig. 1b. Finer grained tuning curves obtained for three typical cells exhibited extremely tight tuning, centred precisely on 0° azimuth with mean half-width at half height values of 3.3° and mean above baseline widths of 16°. Other neurons in this area responded best to stimuli moving in any direction (two-dimensional) through their receptive fields (motion cells; 47%), or preferred some two-dimensional directions over others (directional cells; 36.5%). For the monocular looming-sensitive cells, the directly approaching soccer-ball pattern elicited the same strong response over the entire receptive field, as long as it was on a collision course with the bird's head. Motion of the stimulus in other directions in simulated three-dimensional space produced very little or no increase in the cells' firing rate. The expanding flow pattern produced by the spherical stimulus could have been produced either by the motion of the ball toward the bird, or by the motion of the bird towards the ball. We therefore presented a computer-generated 'whole field' expanding chequerboard flow pattern which filled the whole tangent screen, simulating the approach of the bird toward a large stationary surface. This stimulus pattern produced minimal responses in the looming-sensitive cells. Other controls such as changes in illumination and the presentation of auditory

and somatosensory stimuli also failed to trigger these cells, and their response selectivity for looming stimuli was unaffected by contrast reversal. From these results we conclude that the looming-sensitive neurons are selectively responsive to object motion in depth.

Other experimental observations showed that looming-sensitive neurons were not only selective for object motion in depth, but also gave their maximum response at a constant time before the stimulus would contact the bird, signalling time to collision (Fig. 2). The range of values for time to collision obtained from the sample of 24 looming cells was 800–1,400 ms, with a modal value of 1,000 ms. The neurons exhibited a nearly identical visual response to different sizes of looming stimuli approaching along identical paths at a constant velocity (Fig. 2a), indicating that they are not triggered by the stimulus reaching a critical visual angle. Moreover, when the speed of the stimulus was systematically varied and the path and size kept constant, the cells' response onset and peak firing always occurred at constant times before the stimuli would have contacted the animal (Fig. 2b). Thus, initiation of firing must occur at more distant positions along the stimulus path as velocity is increased, providing strong evidence that these neurons are indeed signalling time to collision. As we have been unable to find neurons with similar characteristics in the optic tectum, which provides the main input to the nucleus rotundus, we conclude that the response properties of these looming cells probably result from patterns of neural connectivity in the rotundus.

It is generally acknowledged that natural or artificial threatening stimuli do not only elicit immediate overt defensive or avoidance behaviour, but also generate autonomic changes including a rapid increase in heart rate and blood pressure^{6,15}. To study these effects we recorded activity from seven looming-sensitive rotundal cells in six birds, while simultaneously recording their heart rate and the muscle activity (electromyograms) from their large pectoralis flight muscle (Fig. 3). The results show a tight correlation between the activity of the rotundal looming-sensitive cells, and the electromyogram and heart rate measurements. When the stimulus was on a collision course

FIG. 2 Peri-stimulus time histograms for single looming-sensitive rotundal cell with a series of soccer-ball pattern stimuli of varying sizes (a), and velocities (b) swept along the direct collision course path toward the bird. Responses are the sum of five sweeps and are referenced to time zero, which is the time when the stimulus would contact the bird. The response remains invariant in amplitude and timing over substantial changes in size and velocity. The simulated path was 15 m long and the simulated size varied from 10 cm to 50 cm in diameter. Velocity in a, 300 cm s⁻¹; size in b, 30 cm.

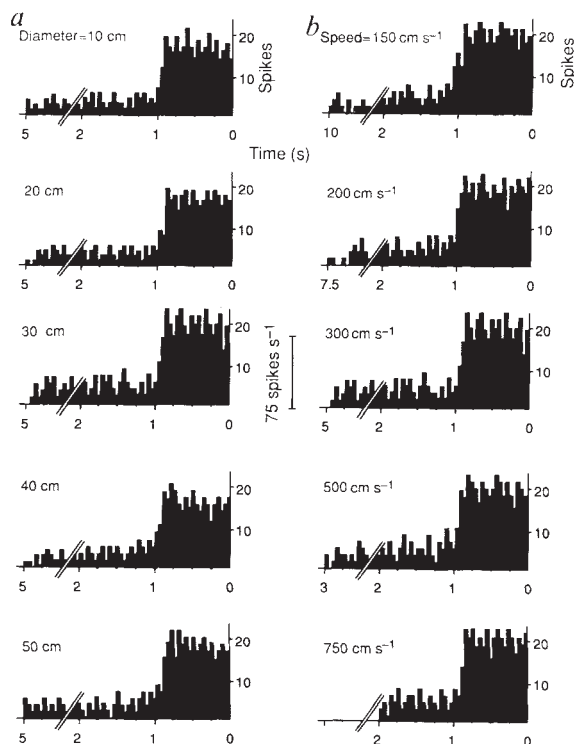
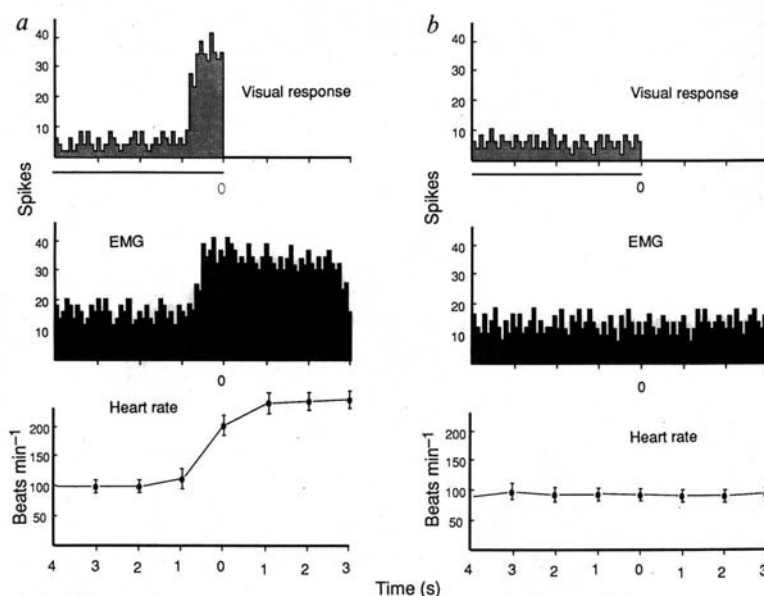


FIG. 3 Heart rate and pectoralis muscle electromyograms (EMGs), recorded concurrently with response rate from a single looming-sensitive rotundal neuron. *a*, Birds were surgically anaesthetized and very fine insulated stainless steel bipolar percutaneous electrodes were inserted into the pectoralis muscle¹⁶. After isolation of a looming-sensitive cell, the birds were allowed to recover from anesthesia just sufficiently to obtain EMGs from the muscle but not overt motion of the bird. No eye movements were discernible. Immediately following the EMG and heart rate recordings, birds were re-anaesthetized to the original level. The looming cell begins firing first, then the muscle response occurs and then heart rate increases dramatically as the soccer-ball looms toward the bird. *b*, No responses occur when the ball moves along the same path but in the opposite direction, directly away from the bird. Bars under the visual response histograms indicate the duration of the visual stimulus. Data collection for the looming-sensitive neuron was terminated with stimulus offset. The neuronal data and EMGs represent summed activity over 5 sweeps of the stimulus whereas the heart rate data represent the means and standard errors for the same 5 sweeps. Simulated size of stimulus was 30 cm, path length 15 m, and velocity 375 cm s⁻¹.



with the bird, sharp increases occurred in firing rate in looming-sensitive cells as well as in heart rate and muscle activity. When the stimulus path deviated from the collision course by as little as 5°, indicating a near miss to either side of the bird, there was a marked reduction in all three responses. The peak muscle activity always occurred 50–100 ms after the onset of the looming-sensitive cell's response, and the heart rate increased about threefold when the stimulus was on a collision trajectory (Fig. 3a). When the stimulus was moving directly away from the animal all three responses remained at baseline levels (Fig. 3b).

For many animal species, including humans and monkeys, responses to a rapidly expanding (looming) image are very similar to those elicited by a real approaching object^{1–6}. It follows that neural mechanisms must extract path and velocity information from dynamic retinal images (including motion in depth) to allow these behaviours to occur. There is considerable evidence that the tectofugal pathway is specialized to respond to the spatial position and motion characteristics of moving objects^{13,14} but ignores self-induced visual motion. The looming-sensitive neurons described here seem well suited to provide information about time to collision of approaching objects. We cannot conclude that rotundal looming-sensitive cells are directly responsible for providing information specifically used to initiate and control avoidance behaviour elicited by such stimuli as there may be other visual structures with similar processing capabilities, but the nucleus rotundus is a strong candidate for this functional role. □

New mammalian chloride channel identified by expression cloning

Markus Paulmichl*, Yi Li†, Kevin Wickman, Michael Ackerman, Ernest Peralta† & David Clapham‡

Department of Pharmacology, Mayo Foundation, Rochester, Minnesota 55905, USA

† Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02115, USA

ION channels selectively permeable to chloride ions regulate cell functions as diverse as excitability and control of cell volume^{1–5}. Using expression cloning techniques, a complementary DNA from an epithelial cell line has been isolated, sequenced and its putative structure examined by site-directed mutagenesis. This cDNA, encoding a 235-amino-acid protein, gave rise to a chloride-selective outward current when expressed in *Xenopus* oocytes. The expressed, outwardly rectifying chloride current was calcium-insensitive and was blocked by nucleotides applied to the cell surface. Mutation of a putative nucleotide-binding site resulted in loss of nucleotide block but incurred dependence on extracellular calcium concentration. The unusual sequence of this putative channel protein suggests a new class of ion channels not related to other previously cloned chloride channels^{6–11}.

We used an expression cloning strategy using the Madin Darby canine kidney (MDCK) epithelial cell line as a source of messenger RNA. MDCK cells contain chloride channels dependent on cyclic AMP, calcium, volume and voltage^{12–15}. *Xenopus* oocytes were injected with water or with polyadenylated mRNA isolated from MDCK cells. Injected oocytes were then assayed for chloride channel activity by two-electrode voltage clamp¹⁶ (Fig. 1). Oocytes injected with native MDCK mRNA (50 nl at 1 ng nl⁻¹) displayed a large chloride-selective current.

A cDNA library in the plasmid vector pcDNA I was constructed (see legend to Fig. 1). RNA transcribed *in vitro* was prepared from pools of 25,000 independent cDNA clones, injected into oocytes and assayed for the expression of activity. A positive

Received 13 August; accepted 20 December 1991.

- Schiff, W. *Psychol. Monogr.* **79**, No. 11, (1965).
- Hayes, W. N. & Saif, E. I. *Anim. Behav.* **15**, 102–106 (1967).
- Schiff, W., Caviness, J. A. & Gibson, J. J. *Science* **136**, 982–983 (1962).
- Swanson, M. T. & Gogel, W. C. *Percept. Psychophys.* **39**, 309–326 (1986).
- Martinoia, C. & Delius, J. D. *Biol. Cybern.* **63**, 127–134 (1990).
- Dean, P., Redgrave, P. & Westby, G. W. M. *Trends Neurosci.* **12**, 137–147 (1989).
- Gibson, J. J. *The Ecological Approach to Visual Perception* (Houghton Mifflin, Boston, 1979).
- Cogshall, J. C. *J. exp. Biol.* **57**, 401–413 (1972).
- Borst, A. & Bahde, S. *Naturwissenschaften* **75**, 265–267 (1988).
- Lee, D. N. & Reddish, P. E. *Nature* **293**, 293–294 (1981).
- Wagner, H. *Nature* **297**, 147–148 (1982).
- Revzin, A. M. *Brain Behav. Evol.* **3**, 195–204 (1970).
- Frost, B. J., Wylie, D. R. & Wang, Y.-C. *Vision Res.* **30**, 1677–1688 (1990).
- Frost, B. J. & Nakayama, K. *Science* **220**, 744–745 (1983).
- Hilton, S. H. *J. exp. Biol.* **100**, 158–174 (1982).
- Loeb, G. E. & Gans, C. *Electromyography for Experimentalists* (Chicago University Press, 1986).

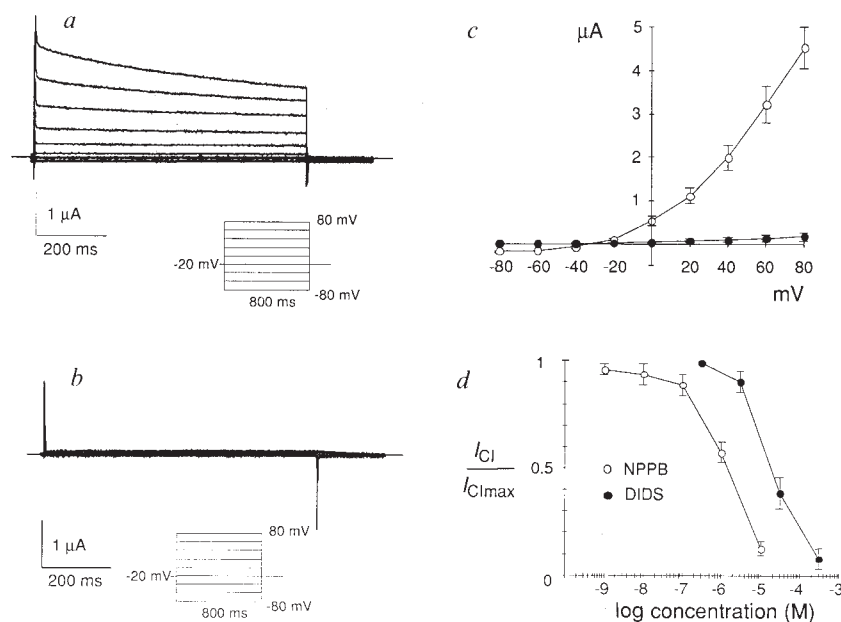
ACKNOWLEDGEMENTS. We thank P. Dodwell, S. Lederman, B. Morgan, M. Robertson, S. Shaver, and D. Wylie for their comments, and T. Kripalani and P. Fearing for computing support. Y.-C.W. was supported by a scholarship from the MRC of Canada. This research was supported by an NSERC grant to B.J.F.

* Present address: Department of Physiology, Fritz-Preglstrasse 3, University of Innsbruck, A-6010 Innsbruck, Austria.

‡ To whom correspondence should be addressed.

FIG. 1 Electrophysiological recordings of I_{Cl} . **a**, Two-electrode voltage clamp recording from a *Xenopus laevis* oocyte injected with mRNA (~50 ng per oocyte) made from pI_{Cl} . **b**, In control oocytes injected with 50 nl water, voltage-activated peak currents averaged 128 ± 22 nA (+40 mV, $n=43$). Holding potential, -20 mV (near E_{Cl}). Voltage steps (800 ms) were made from -80 mV to +80 mV in 20 mV steps. Bathing solutions contained (in mM): 96 NaCl, 2 KCl, 1 MgCl₂, 1.8 CaCl₂, 5 HEPES, pH 7.5. Native, control *Xenopus* oocytes express a calcium-sensitive, relatively voltage-insensitive, chloride current^{17,24} activated by intracellular calcium release from intracellular stores^{25,26}. The calcium-activated chloride current is readily distinguished from the expressed voltage-activated currents we describe. Application of the calcium ionophore A23187 (10 μ M) or injection of calcium in control oocytes (voltage step to +40 mV) resulted in $I_{Ca,Cl}$ averaging, $1,652 \pm 241$ nA ($n=6$) and $2,740 \pm 700$ nA ($n=5$), respectively, but with time courses of activation (τ) at least two orders of magnitude slower than I_{Cl} ($\tau_{A23187} = 176 \pm 16$ ms, $n=6$; $\tau_{Ca} = 175 \pm 15$ ms, $n=5$; compared with <1 ms for I_{Cl} ; not shown). $I_{Ca,Cl}$ was not significantly activated by voltage alone and did not interfere with the measurement of the expressed chloride current. **c**, Current-voltage relationship from control (●) and mRNA-injected oocytes (○; $n=14 \pm$ s.e.m.; the amplitudes of the currents were measured 50 ms after stepping to the different voltages). **d**, Block of I_{Cl} after voltage clamp step to +40 mV (measurements were taken 30–45 s after adding the blockers). The chloride-channel blockers, DIDS and NPPB blocked the current in a dose-dependent manner. DIDS blockade was only partially reversible whereas NPPB reversed fully.

METHODS. Experiments were done using a two-electrode voltage clamp (Turbo TEC 01C, NPI Instruments, Tamm, Germany). Current records were filtered at 300 Hz (8 pole Bessel filter); temperature, 22 °C; all recordings were leak subtracted. In experiments in which extracellular K⁺ concentration was changed, KCl was increased and NaCl decreased accordingly. Extracellular Cl⁻ was decreased by substituting NaCl by sodium gluconate (extracel-



lular free Ca²⁺ concentration monitored by a calcium electrode). MDCK cells were cultured as described¹³. RNA was prepared from MDCK cells using standard techniques²⁷. Double-strand cDNA was prepared with oligo(dT)²⁷ as the first strand synthesis primer; synthetic *Bst*XI adaptors were added to double-strand cDNA. The resulting cDNA was selected for size greater than 1.0 kb and ligated into pcDNA1 (Invitrogen, San Diego). The plasmid cDNA library contained ~500,000 clones. Plasmid DNA was isolated from cDNA pools²⁷. Plasmids were linearized by digestion with *Not*I and transcribed using T7 RNA polymerase. RNA products were concentrated by ethanol precipitation and dissolved in distilled water for oocyte injection. Stage V and VI oocytes were mechanically defolliculated and injected with mRNA within 24 h of surgical removal. Transcripts (50–100 ng per oocyte) were injected in a 50-nl bolus and the oocytes stored for 2 days in supplemented L-15 medium at 19 °C (ref. 24).

```

-71 GGCAGGCGAGCGGCTGGGTCGCCGCGCGGCGAGCTCGCTCCCGGCGCT -23
TGTCTCCGCGCTCCCTGCTGCA ATG AGC TTC CTC AAA AGT TTC CCG CGG 27
M S F L K S F P P 3
CCC GGC TCG GCC GAG GGG CTT CGG CAG CAG CAG CCC GAG ACC GAG 72
P G S A E G L R Q Q Q P E T T E 24
GCC GTG CTG AAC GGC AAG GGC CTC GGC ACC GGC ACG CTC TAC ATC 117
A V L N G K G L G T G T L Y I 39
GCG GAG AGC CGC CTG TCT TGG TTA GAT GGC TCT GGA TTA GGA TTC 162
A E S R L S W L D G G* L G* F 54
TCA CTG GAA TAC CCC ACC ATT AGC TTG CAT GCG GTG TCC AGG GAC 207
S L E Y P T I S L H A V S R D 69
CTA AAT GCC TAT CCA CGA GAG CAT TTG TAT GTT GTG GTG AAT GCT 252
L N A Y P R E H L Y V V V N A 84
AAA TTT GGA GAA GAA TCA AAA GAA TCT GTT GCT GAA GAA GAA GAC 297
L F G E E S K E S V A E E E D 99
AGT GAT GAT GAT GTT GAA CCT ATT GCT GAA TTT AGA TTT GTG CCT 342
S D D D V E P I A E F R F V P 114
AGT GAT AAA TCA GCA TTG GAG GCA ATG TTC ACT GCA ATG TGT GAA 387
S D K S A L E A M F T A M C E 129
TGC CAG GCT TTG CAT CCT GAT CCT GAG GAT GAA GAT TCA GAT GAT 432
C Q A L H P D P D P E D E D S D D 144
TAC GAT GGA GAA GAA TAT GAT GTG GAA GCA CAT GAA CAA GGG CAG 477
Y D G E E Y D V E A H E Q G Q 159
GGG GAC ATC CCT ACA TTT TAT ACC TAT GAA GAA GGA TTA TCC CAT 522
G D I P T F Y T Y E E G L S H 174
TTA ACA GCA GAA GGC CAA GCC ACC TTG GAG AGA TTA GAA GGA ATG 567
L T A E G Q A T L E R L E G M 189
CTT TCT CAG TCT GTG AGC AGC CAA TAT AAC ATG GCT GGA GTC CGG 612
L S Q S V S S Q Y N M A G V R 204
ACA GAA GAT TCA ACC AGA GAT TAT GAA GAT GGG ATG GAG GTA GAC 657
T E D S T R D Y E D G M E V D 219

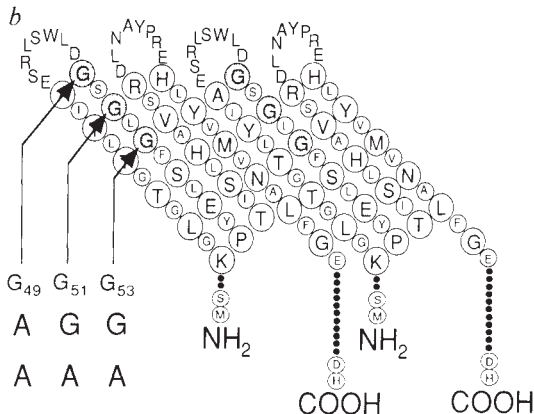
```

```

ACT ACA CCA ACA GTT GCT GGA CAG TTT GAG GAT GCA GAT GTT GAT 702
T T P T V A G Q F E D A D V D 234
CAC TGAAGATGATTATGCTGCTGCTGGGATTCTGCTCATACTAGGAAAGGACTAAAG 760
H 235
AACTTGGTGCTCTTCTACTGTGGAGTGGGTGTTGATGAAAGCTTTTTCTCTCCAA 819
AATTTACCTGAACCACTTCTTTTGGAGACAGCTATATCTAGAGACAAAAGTTGTC 878
ACCAGCAGAAAGAACTATGACCTTTATTAACAAGGGTGAATTAACCTAACCAAGAGGGT 937
ATTGTAGTTTATCATCTACCCAAACCTTTTGTATAGGATTCCTCTGAGTAGGCCCTA 996
TAGTTCCTGCTTCACTATATGCATCTCTGTAACCTGGCAGGCTATGTGGTGAGAAAT 1055
GCGCAGGTGCTTGAAGACATAGGTAGCTGGATGGGAAGGAAGGAGGCTGAATCTAAGA 1114
TATATCCTTTAGAGCCTAAGGATTCCTTACCCCTTAGAAGAAAGACAAAGAGATCTC 1173
AGAGAAATCCTGGCCACTGCTTGTCTTACAGGAGGATTAGGGTCTTTTACTAATTTGG 1232
GATCCAGGTTAATTCATTAGGCATCTTTAGAGAGCCTCTGGAAATATCCTAGCCACA 1291
TGAAATACCACTAATATGGATGCTTCTGATTAGCTTTTACTCTTTCCACCATGTTTCACC 1350
AGTCTGTTTGTATAGCTCTTAGGTTATAGCTTTAATTTCTAACCTCTTAGGTCAG 1409
TTTCCATAACAAATGAATGGGATGAAGTCCTACAGTACTTCAATGATATCTGTTT 1468
GTCTTTGTAATGATTAAACAATAAATCTAGGTTTCTAT 1508

```

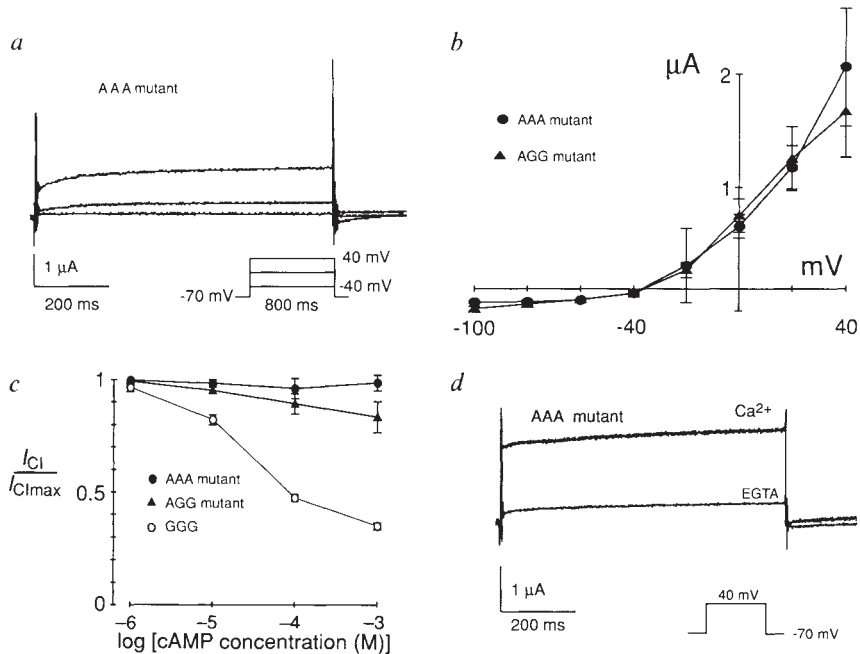
FIG. 2 Nucleotide sequence and predicted amino-acid sequence of cloned cDNA encoding I_{Cl} . Nucleotides are numbered in the 5' to 3' direction, starting at the first ATG coding for the initiating methionine. The nucleotides upstream of residue 1 are indicated by negative numbers. The deduced amino-acid sequence (one-letter code) is shown below the nucleotide sequence. The amino-acid sequence ends at the nonsense codon TGA at the end of the open reading frame. Both strands of the cDNA insert encoding I_{Cl} were sequenced using T7 DNA polymerase²⁸. Solid bars denote the four putative transmembrane β sheets. The asterisk identifies the consensus nucleotide-binding site. Interestingly, there are no consensus sites for protein kinase A or C phosphorylation in the sequence for I_{Cl} . Search of available databases did not reveal significant similarities with other known proteins.



in <1 ms and did not inactivate significantly at potentials hyperpolarized to 0 mV. At +60 mV, $I_{Cl_{in}}$ inactivated slowly with a time constant of 243 ± 15 ms ($n = 10$; Fig. 1) similar to the swelling-induced chloride current^{3,4}. The slow inactivation did not seem to represent chloride accumulation; when oocytes were injected with less (~ 5 ng) p $I_{Cl_{in}}$ mRNA to reduce the net current to <1.5 μ A (+40 mV), the current still inactivated with a mean time constant of 275 ms.

in <1 ms and did not inactivate significantly at potentials hyperpolarized to 0 mV. At +60 mV, $I_{Cl_{in}}$ inactivated slowly with a time constant of 243 ± 15 ms ($n = 10$; Fig. 1) similar to the swelling-induced chloride current^{3,4}. The slow inactivation did not seem to represent chloride accumulation; when oocytes were injected with less (~ 5 ng) p $I_{Cl_{in}}$ mRNA to reduce the net current to <1.5 μ A (+40 mV), the current still inactivated with a mean time constant of 275 ms.

FIG. 4. Mutant I_{Cin} expressed in oocytes 2 days after injection of mRNA transcript (~50 ng per oocyte). *a*, Representative current traces from oocyte expressing AAA mutant. Holding potential -70 mV. Note that the outward current now slowly activates over time on pulsing from a holding potential of -70 mV to +40 mV. Peak current averaged $2,065 \pm 523$ nA at +40 mV (s.e.m.; $n = 11$; fast time constant < 1 ms and a slow time



11; fast time constant <1 ms and a slow time constant of 100 ± 25 ms at $+40$ mV), whereas current in matched water injected control oocytes averaged 132 ± 17 nA (s.e.m.; $n=19$). *b*, Peak current at 800 ms plotted as a function of test potential for AAA and AGG mutant. *c*, Nucleotide block, expressed as in Fig. 3, for mutant and parental channel transcript. cAMP, one of the most effective external nucleotide blockers, was applied to mutant-expressing oocytes. For mutants AAA and AGG, nucleotide sensitivity was lost. *d*, Low extracellular calcium reduces $I_{\text{Clin-mut}}$. Perfusion with 1 mM EGTA solutions reduced net chloride current (bath calcium decreased to $<0.1 \mu\text{M}$ for ≤ 1 min) measured on stepping from a holding potential of -70 mV to $+40$ mV. Average reduction was $81 \pm 3\%$ ($n=5$, $\pm$ s.e.m.). Note in Fig. 3*b* that the putative calcium binding regions²⁹ are immediately adjacent to the nucleotide-binding site, suggesting that alteration of this site has disturbed a possible calcium-binding site by lowering its affinity for calcium. Calcium bound to these sites could regulate the effectiveness of the channel by altering the net positive charge in the external pore region. This putative

mechanism is distinct from the well known regulation of chloride and potassium channels by intracellular calcium.

The I_{Cln} chloride current was detected in 68% of 117 oocytes injected with p I_{Cln} mRNA (oocytes expressing currents >500 nA in <1 ms). Of 943 control oocytes injected with water, only 36 (3.9%) had a similar outward current. No differences from controls were noted in oocytes injected with antisense RNA ($n=15$) or sense mRNA encoding the channel but lacking 65 N-terminal amino acids ($n=19$).

The outward rectifying current measured in oocytes injected with the p I_{Cln} RNA was chloride-selective, as seen from its dependence on external chloride concentration and sensitivity to a selective chloride-channel blocker. Replacement of 50 mM or 90 mM chloride with gluconate shifted E_{Cl} by 10 mV (expected 15 mV) and 40 ± 2 mV ($n=9$; expected 65 mV), respectively. *Xenopus* oocytes respond rapidly to shifts of chloride gradients by active exchange mechanisms, preventing accurate determinations of E_{Cl} . The reversal potential did not shift when the external potassium concentration was increased from 2 to 50 mM. I_{Cln} was blocked by the nonselective anion transport antagonist DIDS (4,4'-diisothiocyanatostilbene-2,2'-disulphonic acid; 50% inhibitory concentration (IC_{50}), 20 μ M), as well as by the more selective chloride channel blocker NPPB (5-nitro-2-(3-phenylpropylamino)-benzoic acid)¹⁸ ($IC_{50}=1.5 \mu$ M; Fig. 1d). I_{Cln} was not calcium-dependent; I_{Cln} was detected in <100 nM external calcium (1 mM EGTA) or after injecting EGTA (~50 nl of 100 mM solution) into oocytes expressing I_{Cln} . In neither case was a significant change of the expressed chloride current detectable.

The nucleotide sequence of the cDNA in p I_{Cln} is shown in Fig. 2. The longest open reading frame encodes a 235-amino-acid protein (relative molecular mass, 25,933). Hydrophobicity profile analysis revealed no potential transmembrane helices. In our proposed model, four β strands are formed by amino acids 31–87. If the channel is a dimer, then an eight-stranded antiparallel β barrel might form the pore¹⁹, similar to the voltage-dependent anion-selective channel from yeast mitochondria²⁰. In this dimeric model, two putative nucleotide-binding sites (GXGXXG; ref. 21) lie near the external mouth of the channel (Fig. 3b).

Figure 3a shows the nucleotide sensitivity of expressed I_{Cln} . As shown, the degree of reversible inhibition ranged from 0% for UMP to ~70% inhibition by cGMP and cAMP. To test whether the putative nucleotide-binding site near the channel mouth was involved in nucleotide sensitivity of the expressed current, the GXGXXG region was mutated in a series of two experiments involving replacement of all three glycines by alanine (G49.51.53A), and by replacement of the first glycine by alanine (G49A; Fig. 3b). Both mutations of the nucleotide-binding site removed cAMP sensitivity (Fig. 4c).

Mutations of the nucleotide-binding site not only altered nucleotide sensitivity, but also altered the current kinetics to those resembling the calcium-sensitive, slowly activating, outwardly rectifying chloride current^{22,23}. Figure 4a, b shows the expressed G49.51.53A mutant. The current remained chloride-selective and displayed a similar current-voltage relationship as the parent protein. Whereas I_{Cln} was insensitive to changes in external calcium concentrations, depletion of calcium in the bath reduced the expressed mutant current ($I_{Cln-mut}$; Fig. 4d). Northern blot analysis of I_{Cln} detected a single transcript of about 1.8 kilobases (kb) in several cell sources including MDCK cells, rat basophilic leukaemia cells, hamster lung, and undifferentiated PC12 cells.

In summary, an expression cloning strategy allowed identification of a novel chloride channel normally expressed in MDCK cells. Mutation of a putative nucleotide-sensitive binding site altered both the nucleotide sensitivity and kinetics of the expressed current, and conferred sensitivity to external calcium concentration (Fig. 4d). These mutations suggest that the nucleotide- and calcium-binding regions in the I_{Cln} channel may be located on the external side of the membrane. The assumption that the protein is a chloride channel is the simplest conclusion,

but more complex interpretations are feasible. The I_{Cln} channel may not only represent an independent family of chloride channels, but also a new structural motif for ion channels. □

Received 31 October 1991; accepted 3 January 1992.

- Lang, F., Völkl, H. & Häussinger, D. *Comp. Physiol.* **4**, 1–25 (1990).
- Hoffmann, E. K. & Simonsen, L. O. *Physiol. Rev.* **69**, 315–382 (1989).
- Worrell, R. T., Butt, A. G., Cliff, W. H. & Frizzell, R. A. *Am. J. Physiol.* **256**, 1111–1119 (1989).
- Solc, C. K. & Wine, J. J. *Am. J. Physiol.* **261**, 658–674 (1991).
- Bretag, A. H. *Physiol. Rev.* **67**, 618–724 (1987).
- Rich, D. P. *et al. Nature* **347**, 358–363 (1990).
- Drumm, M. L. *et al. Cell* **62**, 1227–1233 (1990).
- Frizzell, R. A. *Trends Neurosci.* **10**, 190–193 (1987).
- Quinton, P. M. *FASEB J.* **4**, 2709–2717 (1990).
- Anderson, M. P. *et al. Science* **253**, 202–205 (1991).
- Jentsch, T. J., Steinmeyer, K. & Schwartz, G. *Nature* **348**, 510–514 (1990).
- Kolb, H. A., Brown, C. & Murer, H. *Pflügers Arch.* **403**, 262–265 (1985).
- Paulmichl, M., Wöll, E. & Lang, F. *FEBS Lett.* **234**, 263–266 (1988).
- Lang, F., Defregger, M. & Paulmichl, M. *Pflügers Arch.* **407**, 158–162 (1986).
- Paulmichl, M., Friedrich, F., Maly, K. & Lang, F. *Pflügers Arch.* **413**, 456–462 (1989).
- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 1984).
- Dascal, N. *CRC Crit. Rev. Biochem.* **22**, 317–387 (1987).
- Wangemann, P. *et al. Pflügers Arch.* **407**, S141 (1986).
- Guy, H. R. *Biophys. J.* **47**, 61–70 (1985).
- Blachly-Dyson, E., Peng, S., Colombini, M. & Forte, M. *Science* **247**, 1233–1236 (1990).
- Saraste, M., Sibbald, P. R. & Wittinghofer, A. *Trends biochem. Sci.* **15**, 430–434 (1990).
- Cliff, W. H. & Frizzell, R. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4956–4960 (1990).
- Anderson, M. P. & Welsh, M. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6003–6007 (1991).
- Lechleiter, J. *et al. EMBO J.* **9**, 4381–4390 (1990).
- Lechleiter, J., Girard, S., Clapham, D. & Peralta, E. *Nature* **350**, 505–508 (1991).
- Lechleiter, J., Girard, S., Peralta, E. & Clapham, D. *Science* **252**, 123–126 (1991).
- Maniatis, T., Fritsch, E. & Sambrook, J. *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Tabor, S. & Richardson, C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4767–4771 (1987).
- Heizmann, C. W. & Hunziker, W. *Trends biochem. Sci.* **16**, 98–103 (1991).
- Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488–492 (1985).

ACKNOWLEDGEMENTS. M.P. and Y.L. contributed equally to the main finding of this paper, the cloning and expression of I_{Cln} . We thank R. Guy (Laboratory of Mathematical Biology, NIH) for suggesting the β -barrel model, R. Greger (University of Freiburg, Germany) for NPPB, D. Benos (University of Alabama), for information and preliminary data on a putative chloride channel protein and J. Rae for reading the manuscript. This work was supported by an Erwin Schrödinger and Max Kade fellowship to M.P., a Searle Scholars award to E.P., an AHA Established Investigator award to D.C., and NIH grants to E.P. and D.C.

Trembler mouse carries a point mutation in a myelin gene

Ueli Suter*, Andrew A. Welcher*,†, Tayfun Özcelik‡, G. Jackson Snipes*,§, Bela Kosaras||, Uta Francke¶,‡, Susan Billings-Gagliardi#, Richard L. Sidman|| & Eric M. Shooter**

*Departments of Neurobiology, §Neuropathology, †Genetics and Pediatrics and ‡Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA ||Division of Neurogenetics, New England Regional Primate Research Center, Harvard Medical School, Southborough, Massachusetts 01772, USA #Department of Cell Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

THE autosomal dominant trembler mutation¹ (*Tr*), maps to mouse chromosome 11 (ref. 2) and manifests as a Schwann-cell defect³ characterized by severe hypomyelination⁴ and continuing Schwann-cell proliferation throughout life^{5,6}. Affected animals move clumsily and develop tremor and transient seizures at a young age. We have recently described a potentially growth-regulating myelin protein, peripheral myelin protein-22 (PMP-22; refs 7, 8), which is expressed by Schwann cells and found in peripheral myelin. We now report the assignment of the gene for PMP-22 to mouse chromosome 11. Cloning and sequencing of PMP-22 complementary DNAs from inbred *Tr* mice reveals a point mutation that substitutes an aspartic acid residue for a glycine in a putative membrane-associated domain of the PMP-22 protein. Our results

† Present address: Amgen Inc., Thousand Oaks, California 91320, USA.

* To whom correspondence should be addressed.

identify the PMP-22 gene as a likely candidate for the mouse *trembler* locus and will encourage the search for mutations in the corresponding human gene in pedigrees with hypertrophic neuropathies such as Charcot-Marie-Tooth⁹ and Dejerine-Sottas¹⁰ diseases (hereditary motor and sensory neuropathies I and III).

We previously reported the cloning and expression of the myelin protein PMP-22/SR13 (refs 7, 8, 11). PMP-22 protein expression is restricted to the peripheral nervous system (PNS), whereas another major myelin protein, PLP, is restricted to the central nervous system (CNS). Both PMP-22 and PLP are predicted to be integral membrane proteins with four membrane-associated domains^{12,13}. Mutations in PLP were identified as the primary defect in hereditary CNS myelination diseases of mouse, dog and human (reviewed in ref. 13). By analogy, we considered that mutations in the PMP-22 protein could be responsible for neuropathies specific to the PNS.

As a first step in finding candidate mutations that could involve the PMP-22 gene, we mapped its chromosomal location in the mouse. Amplification by polymerase chain reaction (PCR) of DNA from a standard Chinese hamster × mouse somatic cell hybrid panel (ref. 14) using PMP-22-specific PCR primers did not yield specific products. These hybrid lines contain overlapping sets of all mouse chromosomes except for chromosome 11. When the rat × mouse hybrid cell line RTM9, which contains only two mouse chromosomes, 11 and 13, fused together, was found to be positive for mouse-specific PCR-*AccI* fragments (Fig. 1), the assignment of PMP-22 to chromosome 11 was confirmed.

Two classes of PMP-22 cDNA were isolated by cloning specific PCR products from sciatic nerve RNA of behaviourally and histologically characterized heterozygous (*Tr/+*) mice (Fig. 2). From DNA sequencing, one group of cDNAs represented the PMP-22 wild-type allele, and the second population of cDNAs contained a single non-silent nucleotide exchange within the PMP-22 coding region (Fig. 3a). The detected transition mutation from guanine (G) to adenine (A) leads to the substitution of an aspartic acid residue for a glycine in the PMP-22 protein (Fig. 4a). Direct sequencing of products of two independent PCR amplifications on reverse-transcribed sciatic nerve RNA confirmed the presence of both PMP-22 alleles in heterozygous *Tr* mice (Fig. 3a). The G-to-A base change introduces a new *EcoRV* restriction endonuclease cleavage site in the mutated PMP-22 gene. This finding was exploited in a double-blind analysis of three *Tr/+* mice and three unaffected wild-type

+/+ animals from the same litter, born to a *Tr/+* female mated to a (C57BL/6J × C3HeB/FeJ)*F₁* *+/+* male at the N7 generation of outcrosses to the wild-type *F₁* hybrid stock. Using direct amplification of genomic DNA followed by *EcoRV* digestion, the mutated and wild-type PMP-22 alleles were found in all three *Tr/+* mice; the three normal littermates carried only the wild-type allele (Fig. 3b). As the mice were used at the N7 generation of inbreeding, we conclude that the transition mutation in the PMP-22 gene maps with high probability to the immediate vicinity of the *Tr* locus on chromosome 11.

The predicted effect of a mutation in the PMP-22 protein is consistent with the phenotype of the *Tr* mouse and recent studies on PMP-22 (ref. 11). First, the PMP-22 protein is expressed predominantly by Schwann cells of the PNS and is not detectable in the CNS. In agreement with this, the *Tr* mutation is manifested as a Schwann-cell defect in the PNS with no known consequences in the CNS. Second, the PMP-22 protein is localized in the compact PNS myelin sheath, consistent with the markedly reduced myelination of PNS axons found in *Tr* mutants (Fig. 2). Third, PMP-22 messenger RNA expression is implicated in cellular growth arrest¹². If it is, then the abnormality of PMP-22

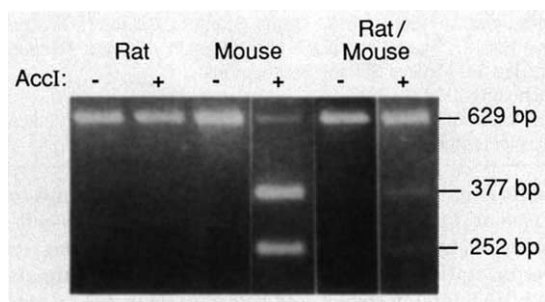


FIG. 1 Mapping of the mouse PMP-22 gene to chromosome 11. DNA isolated from a mapping panel of hybrid cell lines was analysed by PCR (reaction for 1 min at 94 °C, 1 min at 55 °C, 1 min at 72 °C; 28 cycles) using two oligonucleotide primers spanning nucleotides 852 to 868 (5'-TTATGTTGAC-CGTCAGC-3') and 1,483 to 1,467 (5'-AACTGTGGGAGTGACGA-3') of the rat PMP-22 cDNA (ref. 8), identical to positions 919 to 935 and 1,547 to 1,531 (except for nucleotide 1,537) in the corrected mouse *gas3* cDNA (ref. 12; Fig. 4a). Genomic mouse and rat DNA were used as controls in parallel amplifications and *AccI* analyses. Amplification products from the mouse × rat hybrid RTM9 are marked Rat/Mouse. Band sizes are indicated in base pairs (bp).

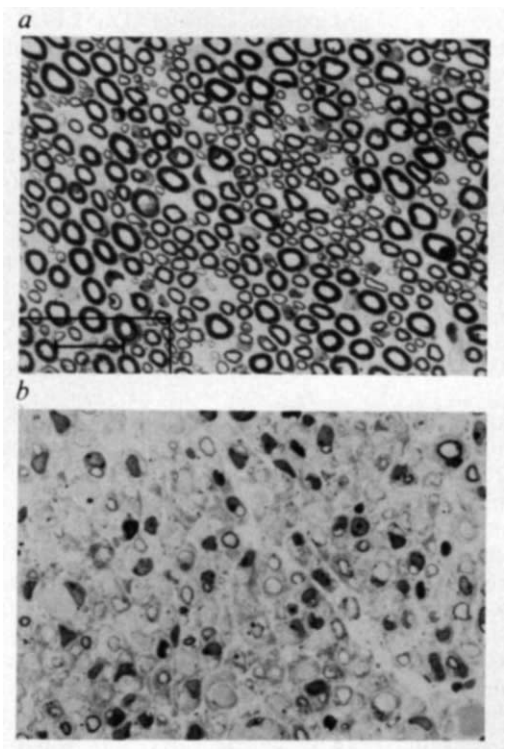


FIG. 2 Transverse sections of sciatic nerves from *+/+* (a) and *Tr/+* (b) 23-day-old littermate mice. Immediately after cervical dislocation, the segment of sciatic nerve in the thigh was dissected out and frozen for RNA extraction (see legend to Fig. 3). The segment of sciatic nerve between the iliac crest and the vertebral column was excised and immersed in 4% paraformaldehyde and 2% glutaraldehyde in 0.1 M phosphate buffer, pH 7.4, at 4 °C. The tissue was washed in the same buffer containing 3% sucrose, and post-fixed in 1% osmium tetroxide in the buffer-sucrose medium for 2 h at 4 °C. The tissue was dehydrated in graded ethanols and embedded in Epon. The 1-μm sections were stained with alkaline toluidine blue. The *+/+* nerve (a) shows densely stained compact myelin sheaths encircling unstained axons of various calibres. The myelin sheath thicknesses are essentially proportional to the calibre of the individual axons they enclose. The *Tr/+* nerve (b) shows relatively few well-myelinated axons, though the axons themselves are similar in calibre to those in the littermate control specimen. Most axons have very thin or absent myelin sheaths, but are contacted by Schwann cells with prominent nuclei and, at times, voluminous cytoplasm. The concentration of nuclei is increased relative to the control, in keeping with previously reported data⁵. Micrographs in a and b are of the same magnification and the scale bar represents 16 μm.

could explain the persistence of Schwann-cell proliferation into adulthood in *Tr* animals. The exact role of PMP-22 in both cellular growth arrest and myelination remains to be determined.

In the process of isolating mouse PMP-22 cDNAs by PCR, we identified a critical frame-shifting sequencing error in the previously reported sequence of the *gas3* cDNA (ref. 12), the putative mouse homologue of rat PMP-22. From our results it is clear that the correct mouse (Fig. 4a), and rat PMP-22 amino-acid sequences⁸ are identical in length and share an amino-acid identity of 97.5%.

The mutation in the PMP-22 gene we report here, associated with the *Tr* phenotype, translates into the replacement of Gly 150 by Asp in the PMP-22 polypeptide (Fig. 4a). This exchange introduces a charged amino-acid residue in the fourth putative transmembrane-spanning domain of the PMP-2 protein (Fig. 4b). Such a mutation is likely to affect the structure and function of an integral membrane protein: the mutated PMP-22 protein may no longer be able to insert correctly into myelin membranes, or the mutation may interfere with the function of a critical domain of the PMP-22 protein. In this context, it is interesting that deleterious mutations in a similar position are found in the PLP gene (reviewed in ref. 13). More structural and functional information on wild-type and mutant PMP-22 proteins is needed to clarify the effects of the putative *Tr* mutation.

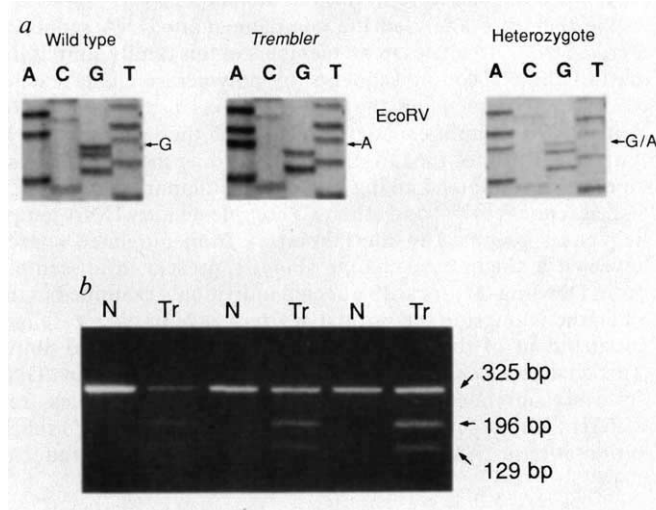


FIG. 3 *a*, Identification of the putative *Tr* allele. Total RNA was isolated from the sciatic nerve of inbred, heterozygous *Tr* mice (a pair of siblings). The RNA was reverse-transcribed and for PCR amplification (for 30 s at 94 °C, 1 min at 50 °C, 1 min at 72 °C; 25 cycles) two oligonucleotide primers spanning positions 91 to 109 (5'-GAGTTTGTGCTGAGGCTA-3') and 832 to 814 (5'-ACAGCAATCCCCACTCAAC-3') in the corrected mouse *gas3* sequence were used¹² (Fig. 4a). PCR products were cloned using a TA cloning kit (Invitrogen) and the entire coding region of 4 independent clones was analysed by dideoxynucleotide DNA sequencing. A G-to-A transition mutation was found in two clones at position 635 (labelled *Trembler*). The sequences of the other two clones were wild-type. To rule out possible PCR errors and to visualize heterozygosity, the products of independent PCR amplifications were directly sequenced (Heterozygote) using a *gas3*-specific oligonucleotide primer (position 521 to 537; 5'-GTGCAGCGGCCACTCTAC-3'). The order of nucleotides on the sequencing gels is given at the top. The respective nucleotides at the critical position 635 are marked with arrows. A novel *EcoRV* restriction endonuclease cleavage site, diagnostic for the putative *Tr* allele, is indicated. *b*, Segregation of the putative *Tr* allele. Genomic DNA was isolated from three heterozygous *Tr* animals and three normal littermates (N). For PCR (30 s at 94 °C, 1 min at 55 °C, 30 s at 72 °C; 30 cycles) two oligonucleotide primers spanning nucleotides 507 to 523 (5'-TCTGTGCGTGATGAGTG-3') and 832 to 814 (5'-ACAGCAATCCCCACTCAAC-3'; ref. 12; Fig. 4a), respectively, were used. PCR products were cut with *EcoRV* and analysed on a 2% agarose gel. The 325-bp DNA fragment represents the wild-type allele; fragments of 129 bp and 196 bp indicate the presence of the *Tr* allele. Results were confirmed by direct sequencing of the PCR products as in *a*.

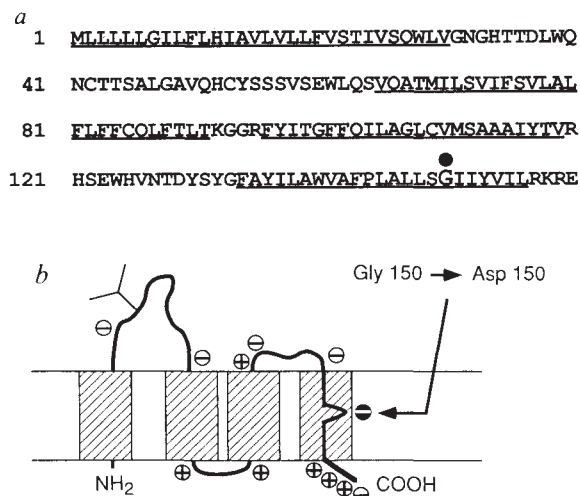


FIG. 4 *a*, Corrected amino-acid sequence (one-letter code) of the mouse PMP-22 protein. The sequence differs from the amino-acid sequence of *gas3* described in ref. 11 in the 25 C-terminal residues owing to a cDNA sequencing error in the original report (a cytosine is missing in the *gas3* cDNA sequence at position 591; ref. 12). The correct nucleotide was initially identified by direct sequencing of PCR products obtained from reverse-transcribed RNA isolated from growth-arrested NIH3T3 fibroblasts (data not shown) and was later confirmed (C. Schneider, personal communication). The site of the putative *Tr* mutation is marked with a dot. Putative membrane-associated domains of the PMP-22 protein are underlined. *b*, Model of the PMP-22 protein structure. The positions of charged amino acids (neutral pH) are indicated. The location of the putative *Tr* mutation is marked with an arrow. The model is based on computer-generated hydrophobicity plots, surface probability and secondary structure predictions¹⁷.

Our data indicate that a non-conservative amino-acid substitution in the PMP-22 protein is most probably responsible for the *Tr* phenotype. Transgenic mice models that either recreate or cure the *Tr* phenotype will be required for final confirmation. Further evidence for the involvement of PMP-22 in peripheral neuropathies will follow when the primary structure is determined of PMP-22 in the *Tr*^J mouse⁵, a mutation allelic to *Tr*. Furthermore, it will be of major interest to examine the status of the PMP-22 gene in selected human inherited PNS neuropathies. The most common form of autosomal dominantly inherited hypertrophic neuropathy, Charcot-Marie-Tooth disease 1a (CMT type 1a), is caused by mutations at the *CMT1A* locus, which has been assigned to region p12-p11.2 of human chromosome 17 (ref. 15). Genes in that region have evolutionarily conserved counterparts on mouse chromosome 11 near the *Tr* locus¹⁶. Therefore we suggest that the human PMP-22 gene will be found on the proximal short arm of chromosome 17, identifying PMP-22 as a candidate gene for the Charcot-Marie-Tooth disorder.

Charcot-Marie-Tooth type 1a is associated with a large DNA duplication of undetermined size which is detected with markers closely linked to *CMT1A* (ref. 9). The relationship of this duplication to the development of the disease is unclear. Searching for PMP-22 gene mutations in CMT type 1a families and mapping its position in relation to the duplicated fragment may clarify the mechanisms involved. □

Received 12 November 1991; accepted 3 January 1992.

1. Falconer, D. S. *Genet.* **50**, 192-201 (1951).
2. Davisson, M. T. & Roderick, T. H. *Cytogenet. Cell Genet.* **22**, 552-564 (1978).
3. Aguayo, A. J., Attiwell, M., Trecarten, J., Perkins, S. & Bray, G. M. *Nature* **265**, 73-75 (1977).
4. Henry, E. W. & Sidman, R. L. *Science* **241**, 344-346 (1988).
5. Henry, E. W., Cowen, J. S. & Sidman, R. L. *J. Neuropathol. exp. Neurol.* **42**, 688-706 (1983).
6. Perkins, C. S., Aguayo, A. J. & Bray, G. M. *Neuropathol. appl. Neurobiol.* **7**, 115-126 (1981).
7. De Leon, M., Welcher, A. A., Suter, U. & Shooter, E. M. *J. neurosci. Res.* **26**, 437-448 (1991).
8. Welcher, A. A., Suter, U., De Leon, M., Snipes, G. J. & Shooter, E. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7195-7199 (1991).

9. Lupski, J. R. *et al. Cell* **66**, 219–232 (1991).
10. Dyck, P. J., Lambert, E. H., Sanders, K. & O'Brien, P. C. *Mayo Clin. Proc.* **46**, 432–436 (1971).
11. Snipes, G. J., Suter, U., Welcher, A. A. & Shooter, E. M. *J. Cell Biol.* (in the press).
12. Manfioletti, G., Ruaro, M. E., Del Sal, G., Philipson, L. & Schneider, C. *Molec. Cell. Biol.* **10**, 2924–2930 (1990).
13. Pham-Dinh, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 7562–7566 (1991).
14. Mönke, M., Cox, D. R., Jackson, I. J., Hogan, B. L. M. & Francke, U. *Cytogenet. Cell. Genet.* **42**, 236–240 (1986).
15. McAlpine, P. J. *et al. Genomics* **7**, 408–415 (1990).
16. Mönke, M. & Francke, U. *J. molec. Evol.* **25**, 134–140 (1987).
17. Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–392 (1984).

ACKNOWLEDGEMENTS. This work was supported by the American Paralysis Association (E.M.S.), the NINDS (S.B.-G., R.L.S. and E.M.S.) and the NCHGR (U.F.). U.S. was supported by the Swiss NSF and the Swiss Academy for Medical Sciences and A.A.W. by a NIH-NRSA fellowship. U.F. is an investigator and T.O. an associate of the Howard Hughes Medical Institute.

A mutation in the conserved helix termination peptide of keratin 5 in hereditary skin blistering

E. B. Lane*, E. L. Rugg*, H. Navsaria†, I. M. Leigh†, A. H. M. Heagerty‡, A. Ishida-Yamamoto§ & R. A. J. Eady§

*CRC Cell Structure Research Group, Cancer Research Campaign Laboratories, Department of Anatomy and Physiology, Medical Sciences Institute, University of Dundee, Dundee DD1 4HN, UK
 †Experimental Dermatology Laboratory and ICRF Skin Tumour Laboratory, 56 Ashfield Street, London E1 2BL, UK
 ‡Department of Dermatology, North Staffordshire Royal Infirmary, Stoke-on-Trent ST4 7LN, UK
 §Department of Cell Pathology, Institute of Dermatology, UMDS, St Thomas's Hospital, London SE1 7EH, UK

IN the hereditary blistering condition epidermolysis bullosa simplex, the skin blisters on trauma following rupture of epidermal basal cells. Clinical variations range from severely incapacitating, especially in early childhood, to mild forms that may not even present clinically. Dowling–Meara epidermolysis bullosa simplex is characterized by clusters of epidermal blisters and keratin clumping in the cytoplasm¹; recent reports describe potentially causal mutations in keratin 14 (refs 2, 3). Here we describe a 'complementary' mutation at the other end of the other keratin expressed by these cells (K5, coexpressed with K14), a change from a Glu to a Gly in the helix termination peptide, detected by altered antibody binding and confirmed by sequencing using the polymerase chain reaction. The two conserved helix boundary peptides are predicted to be essential for filament assembly, and the requirement for two complementary (type I and type II) keratins is absolute. Epidermolysis bullosa simplex diseases demonstrate the function of the keratin cytoskeleton in resisting compaction stresses which otherwise lead to cell lysis.

Keratin samples were analysed from seven family members in a pedigree affected by the Dowling–Meara (herpetiformis) type of epidermolysis bullosa simplex (EBS) over three generations (Fig. 1). When cultured keratinocytes from two patients (siblings IV-2 and IV-3) were analysed by SDS–polyacrylamide gel electrophoresis, the keratin 5 intermediate filament protein (K5) migrated as two bands of relative molecular mass 58,000 and 59,000, instead of the expected single band at 58,000 (ref. 4); however their affected mother (III-2) seemed to express a single band indistinguishable from control samples. These bands were all K5, as confirmed by extensive parallel immunoblotting with monoclonal antibodies to keratins (Fig. 2). Heterozygosity at the K5 locus ($K5_a/K5_b$) is estimated to occur in about 15% of the population, with over 84% of the population homozygous for $K5_b$ (ref. 5). We determined by immunoblotting that the unaffected father is heterozygous ($K5_a/K5_b$). The slower-migrating paternal allotype, $K5_a$, is inherited by the heterozy-

gous affected children (IV-2 and IV-3 in Fig. 1), whereas their second, $K5_b$ -like, allotype comigrates in SDS–polyacrylamide gels with the leading edge of the thick K5 band from the affected mother (Fig. 2). No other individuals with the $K5_a$ allotype were found in this pedigree; all the other affected individuals showed the same (pseudo)homozygotic keratin phenotype as the mother (III-2). Out of a further 18 unrelated individuals sampled, only one (normal) was $K5_a/K5_b$ heterozygous.

An abnormality in the lower keratin 5 band of IV-2 and IV-3 was detected as a specific lack of reactivity with the monoclonal antibody IFA. The IFA antibody normally reacts with all type II keratins and other intermediate filament proteins^{6,7}. Its binding site is in a short peptide at the end of the rod domain⁸ containing the consensus sequence Thr-Tyr-Arg(Lys/Ser)-Leu-Leu-Glu-Gly-Glu, known as the helix termination peptide (HTP in Fig. 4). This peptide is common to all intermediate filament molecules sequenced and it is thought to form a small but specific antiparallel overlap between tetramers at the carboxy-terminal end of the rod domain⁹, which is essential for organized formation of polymers¹⁰. The IFA-negative allotype was not detectable in the other affected family members because the normal IFA-positive $K5_b$ comigrates with the IFA-negative $K5_{DM}$ in the (pseudo)homozygous individuals. Loss of reactivity with the IFA monoclonal antibody strongly suggests a significant change in the protein structure in this subdomain.

We therefore analysed the complementary DNA sequences derived from K5 alleles in six members of this family, amplifying part of the K5 coding sequence by polymerase chain reaction (PCR) and sequencing the PCR products using end-labelled primers. The amplified fragment encoded the last part of helix 1 and nearly all of the tail domain, including the helix termination peptide and its flanking sequences. Comparisons with published sequences^{11,12} and other K5 complementary DNA sample sequences prepared in the laboratory from unrelated sources revealed a single base change (Fig. 3), present in all samples from Dowling–Meara EBS affected individuals examined in this pedigree, alongside the normal K5 sequence in each case, and independent of the $K5_a/K5_b$ polymorphism described above. This change was not seen in the unaffected father, nor in cDNA from six unrelated normal controls, including samples from different EBS types. The identified mutation is an A-to-G substitution at base 1,815 in the sequence of K5 (ref. 11), producing

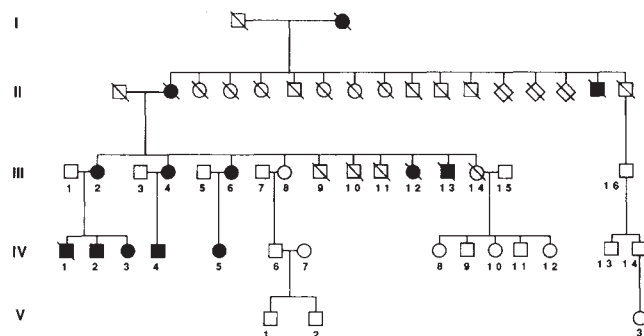


FIG. 1 Pedigree of family affected with epidermolysis bullosa simplex, Dowling–Meara variant, diagnosed by clinical, histopathological and electron microscopical criteria. Of the 56 members documented, 24 are living; the disease is expressed in at least four of the five generations as an autosomal dominant trait. Crossed symbols are deceased; open symbols are unaffected, or not known to be affected; filled symbols are affected, for example, ■, represents male, living, affected; ◆, affected, gender unknown; ○, female, deceased, unaffected, and so on. Skin biopsies (5 mm) were obtained from seven family members for cell culture and protein and mRNA extraction: III-1 (unaffected), III-2 (40 years), III-4 (46 years), III-6 (47 years), IV-2 (19 years), IV-3 (13 years) and IV-5 (24 years). In addition, DNA samples were analysed from III-4 and IV-4 (affected) and III-8, IV-11 and 12 (unaffected).

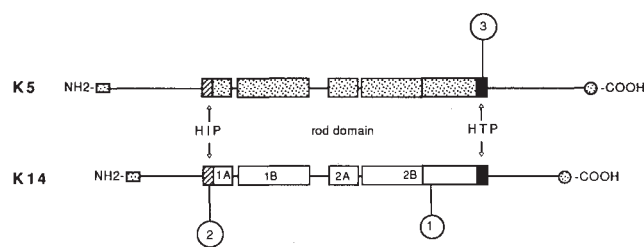


FIG. 4 Diagrammatic representation of keratin 14 (open) and keratin 5 (shaded) to show the position of the mutation described here (3), relative to the recently reported keratin 14 mutations (1, 2) and the recognizable sequences²⁹ of the conserved helix boundary peptides. The α -helical rod domain is the structural subunit of the final 10-nm thick cytoplasmic filament, starting at the helix initiation peptide (HIP; sequence is type-specific), extending through helical domains 1A, 1B, 2A and 2B and their non-helical linker regions to terminate at the helix termination peptide (HTP; in all intermediate filaments); the rod domain is flanked by non-helical amino-terminal head and carboxy-terminal tail domains. The vertical line in helix 2B represents a conserved helix phase reversal.

Cockayne) with chromosomes carrying the major keratin gene clusters², chromosomes 17 (type I keratins) and 12 (type II keratins), respectively. Of these, the chromosome 17 linkage has been confirmed as a mutation in helix 2B of keratin 14 (ref. 16). K14 mutations at a different site were identified independently in two isolated cases of Dowling-Meara EBS³: two point mutations at the same position changed Arg to Cys or Arg to His at an invariant (internal) residue in the conserved helix initiation peptide³ which lies in a polymerization contact region between two associating antiparallel heterodimers⁹. By contrast, our studies have revealed a mutation in the helix termination peptide of keratin 5, that is at the other end, of the other keratin. This supports the possibility¹⁶ that the Weber-Cockayne chromosome 12 linkage may also be due to keratin 5.

The structural pattern common to all intermediate filaments is shown in Fig. 4. Both K5 and K14 are equally important in the structure of keratin filament networks, as keratins are obligate heteropolymers of type I and type II proteins. So are both the conserved helix boundary peptides^{10,17-19} where the EBS mutations have been found. This complementary pattern of mutations in both helix initiation and helix termination peptides in both copolymerizing keratins of epidermal basal cells shows that EBS diseases are probably neither mutation-specific nor even gene-specific, but structure-specific.

The accumulating evidence suggests that many forms of EBS may be associated with mutations in keratin genes. These mutations illustrate the absolute requirement for structural reinforcement by keratin intermediate filaments to maintain tissue and cellular integrity *in vivo*, because they lead to a reduced ability of K5 (here) and K14 to make a stable, resilient, extended keratin filament network. The resulting cell fragility causes rupture of the EBS basal cells on trauma: the keratin filaments are thus functioning, literally, as a cytoskeleton. Other types of cytolytic diseases should now also be scanned for possible intermediate filament alterations, now that the importance of an intermediate filament cytoskeleton can be seen so unequivocally. □

Received 12 November 1991; accepted 3 February 1992.

- Anton-Lamprecht, I. & Schnyder, U. W. *Dermatology* **164**, 221-235 (1982).
- Bonifas, J. M., Rothman, A. L. & Epstein, E. H. *Science* **254**, 1202-1205 (1991).
- Coulombe, P. A. *et al.* *Cell* **66**, 1301-1311 (1991).
- Moll, R., Franke, W. W., Schiller, D. L., Geiger, B. & Krepler, R. *Cell* **31**, 11-24 (1982).
- Mischke, D., Wille, G. & Wild, A. G. *Am. J. hum. Genet.* **46**, 548-552 (1990).
- Pruss, R. M. *et al.* *Cell* **27**, 419-428 (1981).
- Cooper, D., Schermer, A., Pruss, R. & Sun, T.-T. *Differentiation* **28**, 30-35 (1984).
- Magin, T. M., Hatzfeld, M. & Franke, W. W. *EMBO J.* **6**, 2607-2615 (1987).
- Stewart, M., Quinlan, R. A. & Moir, R. D. *J. Cell Biol.* **109**, 225-234 (1989).
- Albers, K. & Fuchs, E. *J. Cell Biol.* **105**, 791-806 (1987).
- Eckert, R. L. & Rorke, E. A. *DNA* **7**, 337-345 (1988).
- Lersch, R., Stellmach, V., Stocks, C., Giudice, G. & Fuchs, E. *Molec. cell. Biol.* **9**, 3685-3697 (1989).

- Cohen, C. & Parry, D. A. D. *Trends Biochem. Sci.* **11**, 245-248 (1986).
- Hatzfeld, M. & Weber, K. *J. Cell. Sci.* **99**, 351-362 (1991).
- Ishida-Yamamoto, A. *et al.* *J. invest. Derm.* **97**, 959-968 (1991).
- Bonifas, J. M., Rothman, A. L. & Epstein, E. *J. invest. Derm.* **96**, 550a (1991).
- Albers, K. & Fuchs, E. *J. Cell Biol.* **108**, 1477-1493 (1989).
- Trevor, R. *New Biol.* **2**, 1004-1014 (1990).
- Vassar, R., Coulombe, P. A., Degenstein, L., Albers, K. & Fuchs, E. *Cell* **64**, 365-380 (1991).
- Rheinwald, J. G. & Green, H. *Cell* **6**, 331-344 (1975).
- Leigh, I. M., Tidman, M. J. & Eady, R. A. J. *Br. J. Derm.* **111**, 527-532 (1984).
- Chapman, S. J. & Eady, R. A. J. *Eur. J. Cell Biol.* **39**, 352-359 (1985).
- Stasiak, P. C., Purkis, P. E., Leigh, I. M. & Lane, E. B. *J. invest. Derm.* **92**, 707-716 (1989).
- Purkis, P. E. *et al.* *J. Cell Sci.* **97**, 39-50 (1990).
- Moll, R., Dhovailly, D. & Sun, T.-T. *Virch. Arch. B* **58**, 105-113 (1988).
- Lane, E. B. & Alexander, C. M. *Semin. Cancer Biol.* **1**, 165-179 (1990).
- Bartek, J. *et al.* *Int. J. Cancer* **46**, 839-844 (1990).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Geisler, N. & Weber, K. *EMBO J.* **1**, 1649-1656 (1982).

ACKNOWLEDGEMENTS. We thank the family members for cooperation and support; J. McGrath, J. Byrne, J. English and A. Smith for clinical help; N. Spurr and M. Boxer for DNA samples; S. Veitch and N. Sultan for technical assistance; T.-T. Sun for antibody AE14; and R. Pruss and colleagues for IFA antibody. This work was supported by the Cancer Research Campaign, the Dysmorphic Epidermolysis Bullosa Research Association and the Dunhill Medical Trust; H.N. is supported by the Billerica Burns Fund.

Glycosyl-phosphatidylinositol linkage as a mechanism for cell-surface expression of immunoglobulin D

Jürgen Wienands & Michael Reth

Max-Planck Institut für Immunbiologie, Stübweg 51, D-7800 Freiburg, Germany

THE B-cell antigen receptor of the IgM and IgD class is a multimeric complex consisting of the membrane-bound form of the immunoglobulin molecule and two other proteins, Ig- α and Ig- β ¹⁻⁸. The Ig- α and Ig- β proteins form a disulphide-linked α/β heterodimer and are encoded by the *mb-1* (refs 9, 10) and *B29* genes^{11,12}, respectively^{13,14}. Surface expression of the membrane-bound IgM molecule requires assembly with the α/β heterodimer^{1,3,4,8}. The IgD molecule, however, can be expressed on the cell surface in an α/β -dependent and -independent form^{8,15}. We show here that in the α/β -independent form the IgD molecule is anchored in the plasma membrane through a glycosyl-phosphatidylinositol linker. In the presence of the α/β heterodimer, most of the otherwise glycosyl-phosphatidylinositol-linked IgD molecule is expressed on the cell surface as transmembrane proteins.

The myeloma line J558L, which does not express the Ig- α protein, is a useful tool to study the control of surface expression of immunoglobulin molecules. The J558L μ m and J558L δ m transfectants do not transport their immunoglobulin molecules to the cell surface^{3,16,17} unless transfected with an *mb-1* expression vector^{4,8,15}. Thus cell-surface expression of the antigen receptor requires assembly of the immunoglobulin molecule with the α/β heterodimer. Nevertheless, we have isolated from the J558L δ m transfectant the myeloma variant J558L δ m2.6 in which the IgD molecule comes to the cell surface in the absence of Ig- α expression³. Furthermore, the IgD molecules from these cells show no detectable association with an α/β heterodimer^{8,15}. We therefore asked whether on the J558L δ m2.6 variant the IgD molecule is anchored in the plasma membrane through a glycosyl-phosphatidylinositol (GPI)-linker (for review see ref. 18).

The J558L δ m2.6 cells were incubated with or without phosphatidylinositol-specific phospholipase C (PLC), an enzyme that can specifically release GPI-linked proteins from the cell surface¹⁹. The cells were then analysed for surface IgD (sIgD) expression with a FACScan (fluorescence-activated cell sorter) (Fig. 1). Treatment of the J558L δ m2.6 cells with PLC in a sucrose

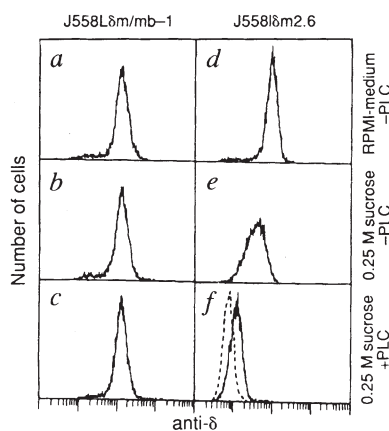


FIG. 1 Release of IgD molecules from the surface of the J558Lδm2.6 cells by phosphatidylinositol specific phospholipase C (PLC). The J558Lδm/mb-1 (a–c) and the J558Lδm2.6 cells (d–f) were incubated for 60 min at 37 °C either with 100 mU ml⁻¹ PLC from *Bacillus cereus* (Boehringer Mannheim) (c, f) or without PLC (a, b, d, e) in the indicated buffers. The 0.25 M sucrose buffer, pH 7.5, has been used because mono- as well as divalent cations strongly inhibit the activity of the PLC¹⁹. The cells were then stained with an FITC-labelled anti-δ antibodies and analysed with a FACScan (Becton Dickinson) J558L cells were also analysed as a negative control (indicated by the dotted line in f). 10⁴ cells were analysed per histogram.

buffer removed more than 90% of the IgD molecules from the cell surface (Fig. 1f). Incubation of the cells in sucrose buffer alone also resulted in some reduction of sIgD molecules (Fig. 1e) compared with incubation in RPMI medium (Fig. 1d). This feature has also been described for other GPI-linked proteins²⁰. As a control we also incubated the *mb-1* transfectant of J558Lδm (J558Lδm/mb-1) with the PLC enzyme and did not see any reduction of the α/β-associated IgD molecules on these cells (Fig. 1a–c). These experiments demonstrate that J558Lδm2.6 cells express GPI-linked IgD molecules on the cell surface and exclude the possibility that PLC treatment results in an

unspecific degradation of surface immunoglobulin (sIg) molecules.

The PLC-mediated release of sIgD molecules into the supernatant was also monitored in a western blot assay (Fig. 2). After PLC treatment, IgD molecules were present only in the supernatant of the J558Lδm2.6 (Fig. 2, lane 8) but not of the J558Lδm/mb-1 cells (Fig. 2, lane 4), although both cell lines express similar amounts of IgD molecules intracellularly (Fig. 2 lanes 1 and 5) and on the cell surface (Fig. 1 a, d). The detection of a distinct δ heavy chain of relative molecular mass of 58,000 (58K) (Fig. 2, lane 8) demonstrates again that the PLC enzyme specifically cleaved the GPI anchor without degrading sIgD molecules.

We next transfected the J558Lδm2.6 cells with the *mb-1* expression vector pLMB1-Hy (ref. 13). The resulting J558Lδm2.6/mb-1 transfectants have two options for transporting the IgD molecules onto the cell surface: as a GPI-linked form or as a transmembrane-bound form in association with an α/β heterodimer. The GPI-linked IgD molecule of the J558Lδm2.6 cells is not associated with an α/β heterodimer (Fig. 3a), whereas the IgD antigen receptor of the J558Lδm2.6/mb-1 transfectants contains a disulphide-linked α/β heterodimer (Fig. 3b). This indicates that the α/β associated form is dominant over the GPI-linked form of IgD. FACScan analysis of the PLC-treated cells reveals two populations (Fig. 4c): one whose sIgD was cleaved, and another whose sIgD was resistant to the enzyme. The assumption that only the latter population expressed the transfected *mb-1* gene was supported by the analysis of an *mb-1*-positive subclone of the J558Lδm2.6/mb-1 transfectant (Fig. 4b, d). The sIgD of this subclone is resistant to PLC treatment (Fig. 4d) and, given the sensitivity of the FACScan, at least 90% of sIgD of J558Lδm2.6 is expressed as transmembrane protein.

The parental J558Lδm myeloma line cannot transport the IgD molecule onto the cell surface³. This is due to the lack of Ig-α expression and to the apparent failure of this line to form a GPI-linker. A cell sorter-derived variant (J558Lδm2.6), however, can express IgD but not IgM molecules on the cell surface³ and we show here that on these cells the IgD is bound through a GPI-anchor. In contrast to IgM, the IgD molecule can be expressed on the B-cell surface either as transmembrane protein^{4,8,15} or as GPI-linked protein. The latter form may also account for the expression of IgD and IgG on the surface of non-lymphoid cells which do not produce the α/β heterodimer⁸.

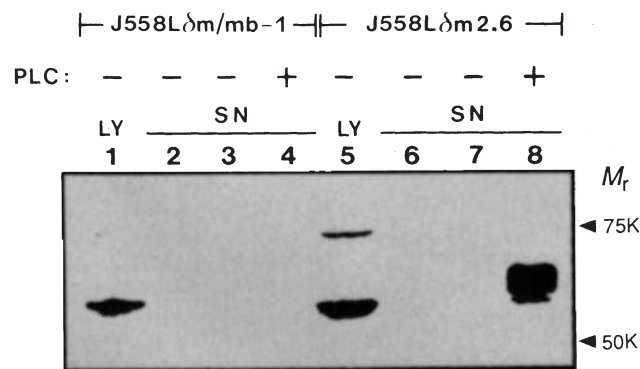


FIG. 2 Detection of IgD molecules in the supernatant of J558Lδm2.6 cells treated with PLC by western blotting. Cells (15 × 10⁶) of the J558Lδm/mb-1 transfectant (lanes 1–4) and the J558Lδm2.6 variant (lanes 5–8) were incubated for 20 min at 37 °C in 200 μl of the following buffers: RPMI medium (lanes 2 and 6); 0.25 M sucrose buffer, pH 7.5, (lanes 3 and 7) or 0.25 M sucrose buffer, pH 7.5, containing 100 mU ml⁻¹ PLC (lanes 4 and 8). Cells were then pelleted and 50 μl from each of the supernatants (SN) were run on a 10% polyacrylamide gel under reducing conditions. As controls, total cell lysates (LY) from the 1.5 × 10⁵ J558Lδm/mb-1 transfectants (lane 1) and from 3 × 10⁵ J558Lδm2.6 cells (lane 5) were run in parallel. After electrophoresis, proteins were blotted onto a nitrocellulose membrane. Proteins were visualized using biotinylated anti-δ antibodies in combination with the avidin-horseradish peroxidase system (ECL-system, Amersham). Arrows indicate the apparent relative molecular mass (*M_r* × 10⁻³).

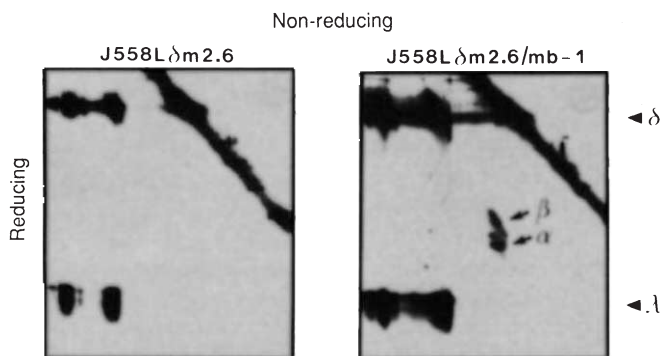


FIG. 3 Transfection by *mb-1* of the J558Lδm2.6 cells leads to association of the IgD molecule with the α/β heterodimer. The IgD antigen receptor of the J558Lδm2.6 variant (a) and its *mb-1* transfectant J558Lδm2.6/mb-1 (b) was affinity-purified¹⁵ from 1% digitonin lysates and analysed by two-dimensional (non-reducing versus reducing 10% polyacrylamide gel) SDS-PAGE. Proteins were stained with silver nitrate²⁸. Stable *mb-1* transfectants were obtained by electroporation²⁹ of the J558Lδm2.6 cells with the linearized pLMB1-Hy vector¹³ and subsequent selection for a hygromycin-resistant population (1.2 mg hygromycin ml⁻¹ culture medium).

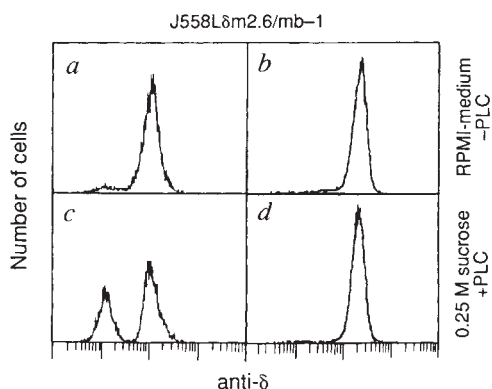


FIG. 4 The mode of surface expression of the IgD molecule is regulated by the presence or absence of the α/β heterodimer. The J558L δ m2.6/mb-1 cells were incubated for 60 min at 37 °C either with 100 μ M ml⁻¹ PLC (c, d) or without PLC (a, b) in the indicated buffers. b, d. One out of 12 mb-1-positive subclones derived from the bulk population shown in a and c. Cells were then stained with an FITC-labelled anti- δ antiserum and analysed with a FACScan (Becton Dickinson). 10⁴ cells were analysed per histogram.

Alternative forms of cell-surface expression have also been found for other receptor molecules (for review see ref. 18). One of the Fc receptors for IgG, Fc γ RIII is expressed on the cell surface either as transmembrane protein in association with the ζ homodimer or as a GPI-linked molecule^{21,22}; the two forms of the Fc γ RIII are encoded by two genes which are differentially expressed²³. In other cases, such as LFA-3, the two different forms of surface expression are generated by an alternative splice^{24,25}. The IgD antigen receptor differs from LFA-3 and Fc γ RIII in that the presence or absence of a second protein, that is the α/β heterodimer, regulates its mode of surface expression.

The transfer of a GPI anchor to a protein in the endoplasmic reticulum seems to be a very rapid process²⁶. Because in the J558L δ m2.6/mb-1 transfectant more than 90% of sIgD is expressed as transmembrane protein in association with the α/β heterodimer it has to be assumed that the assembly of this complex in the endoplasmic reticulum is an equally or more rapid process than the GPI linkage. We think it unlikely that GPI-linked IgD molecules assemble with the α/β heterodimer because binding to the IgM molecule requires the presence of the conserved immunoglobulin transmembrane part⁴. Nevertheless, this possibility has not been completely ruled out.

Our experiments do not exclude the possibility that α/β -producing B cells coexpress the transmembrane form and small amounts of the GPI-linked form of IgD on the surface. As is known from other systems, the two forms differ drastically in their mobility and distribution on the plasma membrane, their internalization route and signalling behaviour²⁷. The different IgD antigen receptors may thus have distinct roles during activation of, or antigen presentation by, B cells. □

Received 2 September 1991; accepted 7 January 1992.

- Hombach, J., Leclercq, L., Radbruch, A., Rajewsky, K. & Reth, M. *EMBO J.* **7**, 3451–3456 (1988).
- Campbell, K. S. & Cambier, J. C. *EMBO J.* **9**, 441–448 (1990).
- Wienands, J., Hombach, J., Radbruch, A., Riestere, C. & Reth, M. *EMBO J.* **9**, 449–455 (1990).
- Hombach, J., Tsubata, T., Leclercq, L., Stappert, H. & Reth, M. *Nature* **343**, 760–762 (1990).
- Parkhouse, R. M. E. *Immunology* **69**, 298–302 (1990).
- Van Noesel, C. J. M., Borst, J., DeVries, E. F. R. & Van Lier, R. A. W. *Eur. J. Immunol.* **20**, 2789–2793 (1990).
- Chen, J., Stall, A. M., Herzenberg, L. A. & Herzenberg, L. A. *EMBO J.* **9**, 2117–2124 (1990).
- Venkitaraman, A. R., Williams, G. T., Dariavach, P. & Neuberger, M. S. *Nature* **352**, 777–781 (1991).
- Sakaguchi, N., Kashiwamura, S. I., Kinoto, M., Thalman, P. & Melchers, F. *EMBO J.* **7**, 3457–3464 (1988).
- Kashiwamura, S. I. *et al. J. Immunol.* **145**, 337–343 (1990).
- Hermanson, G. G., Eisenberg, D., Kincaid, P. W. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6890–6894 (1988).

- Hermanson, G. G., Briskin, M., Sigman, D. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7341–7345 (1989).
- Hombach, J., Lottspeich, F. & Reth, M. *Eur. J. Immunol.* **6**, 2795–2799 (1990).
- Campbell, K. S., Hager, E. J., Friedrich, R. J. & Cambier, J. C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3982–3986 (1991).
- Wienands, J. & Reth, M. *Eur. J. Immunol.* **21**, 2373–2378.
- Sitia, R., Neuberger, M. S. & Milstein, C. *EMBO J.* **6**, 3969–3974 (1987).
- Hombach, J., Sablitzky, F., Rajewsky, K. & Reth, M. *J. exp. Med.* **167**, 652–657 (1988).
- Ferguson, M. A. J. *A. Rev. Biochem.* **57**, 285–320 (1988).
- Ikezawa, H. & Taguchi, R. *Meth. Enzym.* **71**, 731–741 (1981).
- Dhar, A. & Shukla, S. D. *Biochim. biophys. Acta* **1091**, 15–21 (1991).
- Lanier, L. L., Yu, G. & Phillips, J. P. *Nature* **342**, 803–806 (1989).
- Simmons, D. & Seed, B. *Nature* **333**, 568–570 (1988).
- Ravetch, J. V. & Perussia, B. *J. exp. Med.* **170**, 481–497 (1989).
- Seed, B. *Nature* **329**, 840–842 (1987).
- Wallner, B. P., Frey, A. Z., Tizard, R., Mattaliano, R. J. & Hession, C. *J. exp. Med.* **166**, 923–932 (1987).
- Low, M. G. & Saltiel, A. R. *Science* **239**, 268–275 (1988).
- Ferguson, M. A. J. *Curr. Opin. struct. Biol.* **1**, 522–529 (1991).
- Blum, H., Beier, H. & Gross, H. J. *Electrophoresis* **8**, 93–99 (1987).
- Reth, M., Petrac, E., Wiese, P., Lobel, L. & Alt, F. W. *EMBO J.* **6**, 3299–3305 (1987).

ACKNOWLEDGEMENTS. We thank L. Leclercq for reading the manuscript.

A potential fusion peptide and an integrin ligand domain in a protein active in sperm–egg fusion

Carl P. Blobel*, Tyra G. Wolfsberg*,
Christoph W. Turck†, Diana G. Myles‡,
Paul Primakoff‡ & Judith M. White*§

*Departments of Pharmacology and Biochemistry and Biophysics, and

†Howard Hughes Medical Institute and Department of Medicine, University of California San Francisco, California 94143, USA

‡Department of Physiology, University of Connecticut Health Center, Farmington, Connecticut 06030, USA

THE union of sperm and egg is a special membrane fusion event that gives a signal¹ to begin development. We have hypothesized² that proteins mediating cell–cell fusion events resemble viral fusion proteins^{3,4} and have shown that PH-30, a sperm surface protein involved in sperm–egg fusion⁵, shares biochemical characteristics with viral fusion proteins⁶. We report here the complementary DNA and deduced amino-acid sequences of the mature α and β subunits of PH-30. Both are type-I integral membrane glycoproteins. The α subunit contains a putative fusion peptide typical of viral fusion proteins and the β subunit contains a domain related to a family of soluble integrin ligands found in snake venoms. Thus, the PH-30 α/β complex resembles many viral fusion proteins in both its membrane topology and its predicted binding and fusion functions.

Amino-terminal and internal peptide sequences, obtained from PH-30 α and β subunits as they occur on fertilization-competent (mature) sperm⁶, were used to design degenerate oligonucleotide primers for nested polymerase chain reactions (PCR) on guinea-pig testis cDNA. The deduced amino-acid sequences of the PCR products were found to include previously determined peptide sequences. The remainder of the α sequence was generated by a nested rapid amplification of cDNA ends (RACE) protocol⁷ (Fig. 1). The remainder of the β sequence was obtained from a cDNA clone isolated by probing a guinea-pig spermatogenic cell cDNA library⁸ with the β PCR product (Fig. 2).

The nucleotide sequence of mature PH-30 α contains one open reading frame encoding 289 amino acids with one predicted transmembrane domain. Although the glycosylated protein has a relative molecular mass of 60,000 (M_r 60K) on reducing SDS-polyacrylamide gels, the predicted size of unmodified mature PH-30 α from the cDNA sequence is 29,700. Although

§ To whom correspondence should be addressed.

N-linked glycosylation⁶ accounts for part of this difference, additional post-translational modifications, such as *O*-linked glycosylation, or other properties of the molecule must also affect its electrophoretic mobility. The nucleotide sequence of mature PH-30 β contains one open reading frame encoding 353 amino acids, with two potential sites for *N*-linked glycosylation

①	TACTGCACTGGTCAATCTGGAAGTGCCTCTAGACACATACAAACAGATGGTACCCCA	60
	Y C T G Q S G K P C P L D T Y K Q D G T P	20
	TGTAATGAGGGATTCTTCTGTGAAGTGGGTGCACGGACCTGGTATTCAATGCGCA	120
	C N E G F E C G K L I C T D P G I Q C A	40
	ACCTATTTTGGGCATGGTCCAGGCTGCCCCAGATGCATGTTACATACATTGAACAGC	180
	T Y F G H G A R S A P D A C Y T T L N S	60
	ATAGGGAATATATTTGAACTGTGGTCAATCAGGTAATCCGACCACCTATGTTGGTGT	240
	I G N I F G N C G Q S G N P T T Y V G C	80
	TCAGGTGATAGTACAAAGTGTGGAACTCATATGTACAGGATTCTTCAATACCTCCA	300
	S G D S T K C G K L I C T G I S S I P	100
	ATCAGAGCTCTATTTCAGCGATCCAGATCCCTCATGGAGTACTGGTGTGGAGCATT	360
	I R A L F A A I Q I P H G D D W C W S I	120
	AGTAACCTTTGGGATCCTGCGTCTCCCCACAGAGAGCTGTGTCAGCAGGACGCTCT	420
	S N F G D P A S S P T E G A V S A G T S	140
	TGCGCTTCAGGCAAGCGTGTGAATGCCAGTGTCTACTTTCACACTTGACACTGCT	480
	C A S G K A C V N A Q C S T F T L D T A	160
	AACGTAGTCAGCTGAAATGTGTATGAGAATGGAATTGCAACAATTAGGCGACTGC	540
	N C S A A E M C N E N G I C N N L G H C	180
	CACGTGTGAGATGGTTTGGCTCCCCCACTGCAAGCAACAGGAAGTGGAGGTAGTATA	600
	H C G D G F A P P N C K E Q G T G G S I	200
	GATAGTGGTCCCCCTCCCTCTTAGTACACCTACTGCACTCCCTAAACCAACCCAAACG	660
	D S G P P P P S S T P T A P P K P T Q T	220
	ACAAAAGCATCAAGTGAAGTCTAGCATTAAATGGCTGATAATTTGGTAATACTATTA	720
	T K A S S E N L A L I G L I I L V I L L	240
	CTACTTCTAGTTATTTGGTCTATCTGCTTGTATACCTGTAAGAAGCTCCTCCACCA	780
	L L L V I C A I C L G I P A E E A P P	260
	CCAGAAGAGGAAGCAGGGGAACTGGAAGAAGAACAGAACCCAGAACCCGAG	840
	P E E E A G E A G L E E P E P E P E P E	280
	GAGGAGGAAGCAGCAGAGGAGGAAGTAGAGTGATAACTGGTGAAGGGCAAGCCAAA	900
	E E E A A E E E D	289
	TATATCAAAATCTCAAGTGATTACTAGGAATGAGAAGCTCAGGGAAGCAAAAGTTCAGT	960
	GGGAAGTACAGCTCTAATGGCTCTAAGTACGACTATACCTTAAAGACACCACTGT	1020
	AAGAGGACCTATTCTCCATGTTCTCATCTAAGTAATAAATCTTGTACCTAGCAT	1080
	ATTAAACAAAAA	1104

FIG. 1 Nucleotide and deduced amino-acid sequences of mature PH-30 α . Numbered arrows indicate primers used in the initial PCR; the polyadenylation signal AATAAA is heavily underlined. Peptide sequences obtained from purified protein are underlined; where identity is ambiguous, amino acids have a dotted underline. The putative transmembrane domain is boxed and the potential *N*-linked glycosylation site is indicated with an asterisk. METHODS. Mature PH-30 protein was affinity-purified as described⁵ and subunits were separated by SDS-PAGE and electroeluted using a Schleicher and Schuell elutrap. For N-terminal peptide sequence, the eluted protein was reduced and purified by reverse-phase HPLC. For internal peptide sequence, the eluted, reduced protein was cleaved with trypsin before HPLC. Peptides were sequenced by automated Edman degradation using an Applied Biosystems gas phase sequencer. Two peptide sequences (underlined) were used to design four degenerate oligonucleotide primers. Primary PCR on first strand cDNA from guinea-pig testis RNA used primers 1 and 4. The products were diluted and used in secondary PCR with primers 2 and 3. This reaction generated one band of 308 nucleotides which was subcloned, sequenced (Sequenase; US Biochemical) and shown to encode the previously determined peptide sequences. The nested RACE protocol was then carried out as described⁷ to complete the 3' sequence of mature PH-30 α and to confirm the peptide sequence of the mature N-terminus. Amplified DNA fragments were excised from an agarose gel, purified with GeneClean (Bio 101, Inc.), and directly sequenced: 100–500 ng DNA was diluted in 6.75 μ l water, 2 μ l 5 $\times$ reaction buffer and 1.25 μ l of 10 μ M primer, heated for 2 min at 95 $^{\circ}$ C and immediately frozen in dry ice-ethanol; after thawing on ice, DNA was sequenced with Sequenase. Both strands were sequenced with primers spaced at $\sim$ 175 nucleotide intervals; the secondary reaction products from two different primary 3' RACE reactions were identical.

	GAATCCAGCCGCTCTACAGG	21
	P V Y R	-1
①	TCAAACCCGGTCTGCGGCAATAACAGGGTGAACAGGGTGAAGACTGCGACTGCGGATCG	81
	S N F V C G N N R V E Q G E D C D C G S	20
	CAGGAGGAATGCCAAGACACCTGCTGTGATGCTGCTACTGCGAGCTGAAAGTACTTCA	141
	Q E E C Q D T C C D A A T C R L K S T S	40
	CGATGTCTCAAGGGCCCTGTTGAACAGTGTGAGTTCAAACTAAAGGAGAGGTATGC	201
	R C A Q G P C C N Q C E F K T K G E V C	60
	CGAGATCCAGGATGAGTGTGATCTCCCTGAGTACTGCAACGGCTCGTCTGGGGCTTGC	261
	R E S T D E C D L P E Y C N G S S G A C	80
	CAAGAAGACCTTATGTTTATTAATGGGCACAGATGCGCAATGAAGAGTGGATCTGCATG	321
	Q E D L Y V I N G H R C A N E E W I C M	100
	AATGGGAGGTGTCTGTCTGGGAAGGCACATGCCAAGAACATTTGGTACAGAAATGGAA	381
	N G R C L S G K A Q C Q E T F G T E M E	120
	ATGGGTTCAGTAGACTGCTTTGAACAACCTAATACTAAGAATGACATTACTGGAACTGT	441
	M G S V D C F E Q L N T K N D I T G N C	140
	GGTATCTCAGCCCCGGGAATTACAAAGCATGTGGAGTACGAAATGGAGTGTGGA	501
	G I L S P G N Y K A C G A S N W K C G	160
	TTAATCTTCTCTATGATAAGAGCGAAATCTGAGAAACAGGAGCATGACCTTTAT	561
	L I C S Y D K S E I L R N K E G M T I Y	180
	GCCAACTCAGCGGCCATATCTGTGTCAGCATAGAATCTCTGGTCAAGCCCAAGAGT	621
	A N I S G H I C V S I E Y P P G H A K S	200
	GCACTGTATGGTAAAGAGATGGCAGCGTGTGGCCGAGTGGAGTTTGTAGGCAACAA	681
	A L M W V R D G T V C G P S E V C R Q Q	220
	CAGTGTGTATCCAGTTCGTATTGGGATATGACTGCACACCGCCACITGCGATGATCAT	741
	Q C V S S S Y L G Y D C T F A T C S D H	240
	GGTGTATGCAATAACAAAGGCATGTCTGTAATCCCACTTATGTACCTCCAACTGC	801
	G V C N N K R H C H C N P T Y V P P N C	260
	GAGACCAAGATTCGACAAAGCCTGGAGGAGTGTGACAGCGGTAAATCTACGATATGAA	861
	E T Q D S T K P G G S V D S G N L R Y E	280
	CCAATCCCTGAAACGTATTTCGTGAAGTGCTTACCATAACAGTCTAGAAAATGGCCA	921
	P I P E T Y F V E G A Y H T K S R K W P	300
	TTTTCTTGATCATCTCTTTTTCGTATTCTTCGCTACGTTGCTACAGTGTGTA	981
	F F L I I P F F V I F S V L V A T V V K	320
	GTCTATTACCAAAAGAAAAATGGAAACTGAAGATTATGCAATGATGAGAACATTGAA	1041
	V Y Y Q K K K W K T E D Y A N D E N I E	340
	AGCGAGAGTGAACCCAAAGCTCCAAAGTCTCTTCCAAGTAGACTGGCTGGCAAGGATTC	1101
	S E S E P K S S K V S S K	353
	TATGCTGCCACAGTGTGTCAAATAGTTCAAGTCTTCCAGAACAGTATCTTTAGTGA	1161
	TAGAAAAAGTGGAGAAGAAAAATATGCACTACCTTCTTCTGGAAGTCAAAGTATG	1221
	TATCGTGGTTCAGTGCACGAAACTATAGACATTTTCATGTACATGTAAACATACATATA	1281
	TGTCGTATATATGATTCATATAACAGATAATTTATTTGTAAGGAAGGCCTAATTATGAG	1341
	TTTACATTACATTTTCTCATTTTAAAGTTATTTACACCGTGTAGCTAGAGTACAC	1401
	TAACTCTGTAGTAGCATGATATAGAAAAAGTTATTAAGAGGTAATACCATGGGAATT	1461
	CC	1463

FIG. 2 Nucleotide and deduced amino-acid sequences of mature PH-30 β . Numbered arrows correspond to the position of degenerate oligonucleotide primers used in the primary PCR and the polyadenylation signal AATAAA is underlined. Protein sequences of purified peptides are underlined; a dotted underline denotes ambiguous identity. The putative transmembrane domain is boxed and asterisks denote two potential *N*-linked glycosylation sites. METHODS. Peptide sequences were determined as for Fig. 1, but to generate N-terminal peptide sequence, the β subunit was transferred to an immobilized membrane³⁷. To generate internal peptide sequence, the eluted protein was cleaved with cyanogen bromide or V8 protease, electrophoresed, and transferred to immobilized³⁷. Alternatively, eluted protein was digested sequentially with CNBr and trypsin, followed by reverse-phase HPLC. Peptide sequences were used to design degenerate oligonucleotide primers and to confirm the identity of the PH-30 β cDNA clone as in Fig. 1. Primary PCR used primers 1 and 4 on first strand cDNA from guinea-pig testis mRNA which was diluted and used as a template for nested PCR with the primer combinations 1 and 3 or 2 and 3, generating strong bands of 206 and 185 bases, respectively. The 206-base fragment was excised from a polyacrylamide gel, eluted in water, precipitated, sequenced and found to encode PH-30 β peptide sequences. It was labelled with ³²P and used to probe a λ gt11 guinea-pig spermatogenic cell cDNA library⁸ using high stringency hybridization and wash conditions³⁸. One clone, which encoded all of mature PH-30 β plus 4 amino acids of the precursor domain was identified out of 5×10^5 plaques and directly sequenced using PCR-amplified DNA fragments with primers spaced at $\sim$ 200 nucleotide intervals. The sequence was confirmed by sequencing both strands of mature PH-30 β amplified by PCR from first strand guinea-pig testis cDNA.

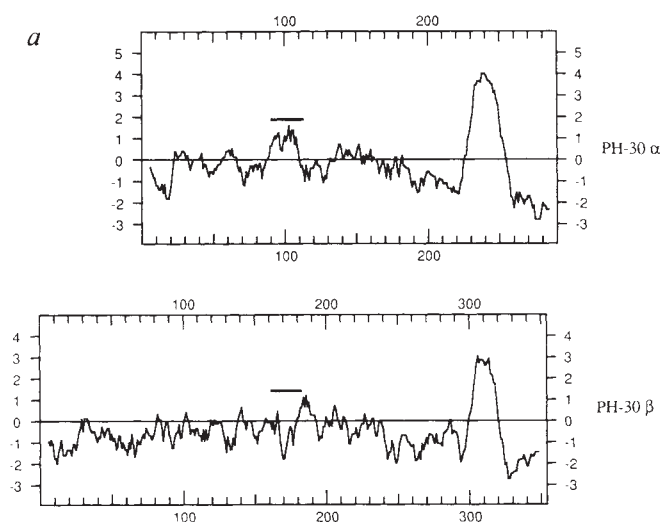
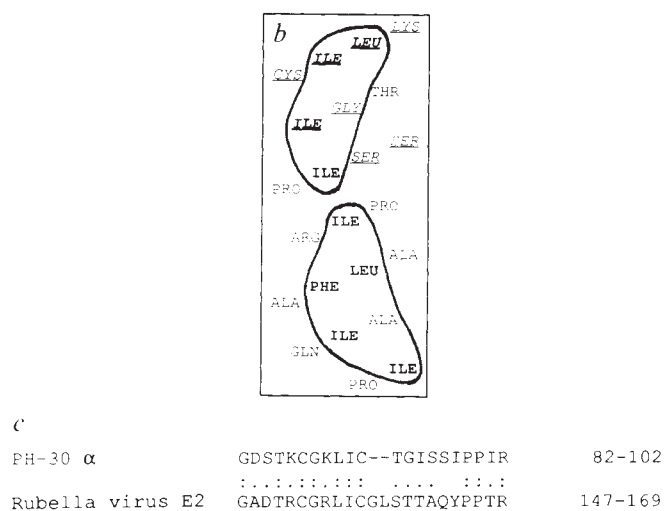


FIG. 3 Putative fusion peptide in mature PH-30 α . a, Hydrophobicity plots of mature PH-30 α and PH-30 β . The region of the putative fusion peptide in α and the corresponding region in β are indicated by bars. The plots were generated according to the coefficients proposed by Kyte and Doolittle using the sequence analysis program DNA Strider 1.1 (ref. 40). b, Residues 90-111 of mature PH-30 α modelled as a sided α helix. Bulky hydrophobic amino acids (Ile, Leu, Met, Phe, Trp, Val) are in bold; residues similar to those in rubella virus E2 (see c) are italicized and underlined. The encircled amino acids define the hydrophobic face. Amino acids were plotted in helix form

and one predicted transmembrane domain. The calculated size of unmodified mature PH-30 β is 39,000, in good agreement with the observed size of glycosylated⁶ PH-30 β on reducing SDS gels⁵ ($M_r = 44K$). The presence of glycosylation sites⁶ and disulphide-linked cysteines⁵ in the N-terminal portions of PH-30 α and β suggests that both are type-I integral membrane proteins with large N-terminal extracellular domains and short cytoplasmic tails.

Most viral fusion proteins contain a 'fusion peptide', either at the N terminus or within the polypeptide, the sequence of which is highly conserved within, but not between, virus families⁴. Fusion peptides are (1) always located in a membrane-anchored subunit, (2) relatively hydrophobic and (3) capable of being modelled as a 'sided' α helix with most of the bulky hydrophobic residues on one face⁴. In at least one case⁹, a helical structure is adopted when the fusion peptide interacts with a membrane surface. The hydrophobic face of the helix may mediate virus-cell membrane interactions that lead to fusion⁴. PH-30 α contains a region (residues 90-111) that fulfills all three of these criteria as an internal fusion peptide (Fig. 3a, top panel, bar; Fig. 3b). Putative internal viral fusion peptides have prolines at or near their centres⁴; similarly, this region has two prolines in the middle of its predicted helix (Fig. 3b) which probably cause a kink, as do prolines in other membrane-interactive α helices^{10,11}. This potential fusion peptide of PH-30 α overlaps a region (residues 82-102; Fig. 3c) that is similar in sequence to a potential fusion peptide (T. G. W. and J. M. W., unpublished data) of the E2 glycoprotein of rubella virus^{12,13}. The significance of this finding is not yet clear¹⁴.

Most viral fusion proteins are responsible for binding, as well as fusion, to target membranes²⁻⁴. The N-terminal 90 amino acids of mature PH-30 β contain a putative integrin-binding 'disintegrin' domain (Fig. 4a) that could function in the membrane-binding step that precedes sperm-egg fusion^{1,2}. Disintegrins such as batistatin¹⁵, barbourin¹⁶, kistrin¹⁷ and echistatin¹⁸ comprise a family of short (49-83 amino acids), soluble and highly conserved platelet aggregation inhibitors from snake venoms which act by competitive inhibition of fibrinogen binding to the integrin gpIb/IIIa (refs 17, 19). The similarity of sperm PH-30 β to these integrin ligands suggests that it



using the program HELO (R. M. Stroud, personal communication). c, Region of similarity between residues 82-102 of mature PH-30 α and residues 147-169 of rubella virus E2 (ref. 13). Double dots indicate identical amino acids; single dots indicate conserved amino acids. The sequences are 44% identical calculated from the program RSS ($z = 7.2$; ref. 39) using 100 random shuffles and a local window of size 10). The similarity was found by searching the NBRF PIR database release 28 using the program FASTA³⁹ with $ktup = 2$ and $PAMFACT = 1$.

binds to a receptor, probably an integrin, on the egg plasma membrane.

Most disintegrins contain an integrin consensus binding sequence RGD (Arg-Gly-Asp; refs 20, 21; Fig. 4a, asterisks). Changes in this sequence alter either the affinity²² or the specificity of binding^{16,19}. Half of the known integrin ligands, however, either do not contain this tripeptide or can interact with integrins using non-RGD sequences²³. Nuclear magnetic resonance analysis revealed that the RGD sequences of kistrin and echistatin are at the tip of a loop, where they are thought to assume a conformation necessary for high-affinity binding^{24,25}. If the disintegrin domain of PH-30 β adopts a similar tertiary structure, the sequence TDE (Thr-Asp-Glu) and an additional (odd-numbered) cysteine would occupy the tip of such a loop (Fig. 4a, asterisks) and might then interact with an integrin on the egg plasma membrane. The disintegrin domain of the longer snake venom protein HR1B (ref. 26) contains the sequence ESE (Glu-Ser-Glu) and an additional (odd-numbered) cysteine in this position (Fig. 4a, asterisks).

Two lines of indirect evidence support the hypothesis that PH-30 β binds to an egg plasma membrane integrin: proteolysis of mouse eggs reduces their fertilization competence^{27,28} and two of the implicated egg surface polypeptides have M_r s of 95K and 150K (ref. 28), similar to α and β integrin subunits; also, micromolar concentrations of RGD-containing peptides inhibit fusion between human or hamster sperm and hamster eggs lacking a zona pellucida²⁹. RGD peptides might compete with PH-30 homologues on these sperm, even if the homologues do not contain an RGD sequence, as RGD peptides can compete for binding of non-RGD integrin ligands^{19,30,31}.

Several aspects of fertilization could be explained if PH-30 β is an integrin ligand. First, although the principal barrier to cross-species fertilization is the zona pellucida^{32,33}, the interaction between species-specific pairs of PH-30 homologues and integrins could explain the additional specificity of sperm-egg plasma membrane interactions³³. Second, sperm-egg fusion requires divalent cations³³, which are also necessary for the function of virtually all integrins²³. Third, signalling events that trigger development accompany sperm-egg fusion^{33,34} and integrins are implicated in signalling processes^{19,35}. Finally, pro-

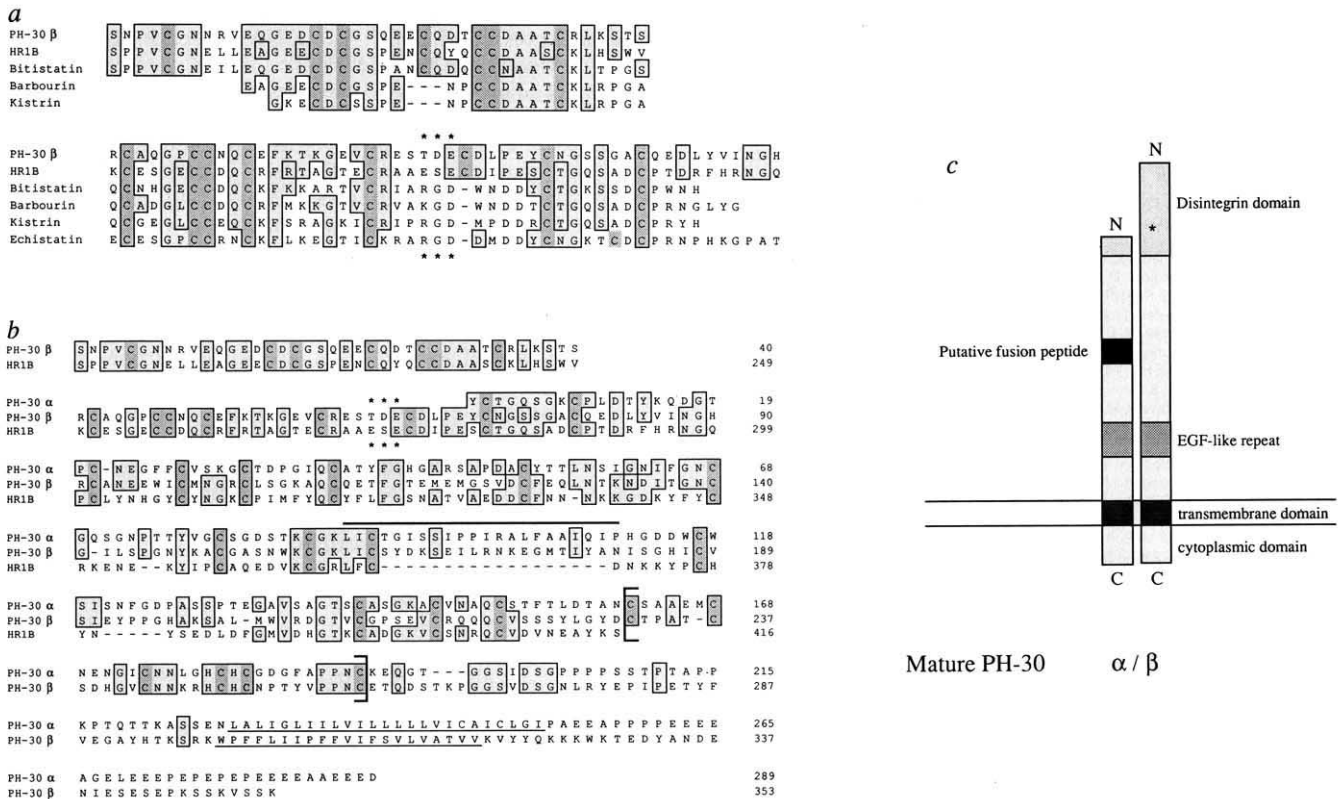


FIG. 4 Similarity of PH-30 β to disintegrins and to PH-30 α . *a*, The N-terminal disintegrin domain of mature PH-30 β and the internal disintegrin domain of the snake venom protein HR1B (ref. 26) are aligned with the complete sequences of 4 representative disintegrins: bitistatin (83 amino acids)¹⁵, barbourin (73 amino acids)¹⁶, kistrin (68 amino acids)¹⁷ and echistatin (50 amino acids)¹⁸. Amino acids identical to those in PH-30 β are boxed, the position of the disintegrin RGD sequence is marked with asterisks and all cysteines are shaded. The similarity between PH-30 β and bitistatin was found by searching the NBRF PIR database as in Fig. 3c. The percentage of identical amino acids in the disintegrin domains compared to that of PH-30 β is 53% for HR1B ($z = 20.0$), 54% for bitistatin ($z = 19.3$), 47% for barbourin ($z = 15.5$), 44% for kistrin ($z = 13.7$), and 36% for echistatin ($z = 9.2$). Statistical comparisons were generated as in Fig. 3c; the alignments shown here were done by hand and differ slightly from those generated by the program. *b*, The complete deduced protein sequences of mature PH-30 α and mature PH-30 β aligned with part of the snake venom protein HR1B (ref. 26) (the upstream metalloprotease domain of HR1B is not shown). Asterisks are in

the same position as in *a*; the EGF-like repeat³⁶ is bracketed; all cysteine residues in the three proteins align and are shaded and identical amino acids are boxed. The putative transmembrane domains of PH-30 α and β are underlined; the cytoplasmic tail of PH-30 α is highly acidic. The N-terminus of α aligns with the C terminus of the disintegrin domain of β , shown in *a*. The putative fusion peptide in α is positioned beneath the bar. Residues 1–92 of PH-30 α are 39% identical to residues 72–163 of PH-30 β ($z = 27.3$), and 34% identical to residues 281–370 of HR1B ($z = 18.2$); residues 72–163 of β are 33% identical to residues 281–370 of HR1B ($z = 20.9$); residues 205–260 of β are 39% identical to residues 135–191 of α ($z = 14.2$); residues 390–416 of HR1B are 61% identical to residues 135–161 of PH-30 α ($z = 8.6$) and 47% identical to residues 205–231 of β ($z = 9.0$). Statistical comparisons were generated as in Fig. 3c; sequences were aligned manually and using computer programs. *c*, Mature PH-30 α and β are depicted as they are oriented in the sperm plasma membrane; the asterisk shows the position of the putative binding loop in PH-30 β .

teolytic processing of PH-30 β correlates with the acquisition of fertilization competence⁶; removal of the precursor domain might expose the integrin ligand domain and thus be required for sperm binding to the egg plasma membrane.

An intriguing finding is that PH-30 α and β seem to be evolutionarily related to each other. Both contain an epidermal growth factor (EGF)-like repeat³⁶ (Fig. 4b) and, when their sequences are aligned from this repeat, they share significant amino-acid identity and every cysteine, the transmembrane domains, and the cytoplasmic tails fall into register (Fig. 4b). Furthermore, all of the cysteines and some other amino acids in α and β align with those of the snake venom protein HR1B (Fig. 4b). HR1B, however, is truncated immediately before the EGF-like repeat and lacks a transmembrane domain (Fig. 4b). The region of HR1B that would correspond to the fusion peptide of PH-30 α is absent (Fig. 4b). In PH-30 β , this region does not model as a sided α helix (T.G.W. and J.M.W., unpublished data). The mature PH-30 α subunit contains (at its N terminus) the C terminus of a disintegrin domain (Fig. 4b and c). Analysis of the precursor domain of PH-30 α revealed that it contains a complete disintegrin domain (data not shown). The disintegrin

domain of the α precursor, which is cleaved earlier in spermatogenesis⁶, might attach spermatogenic cells to Sertoli cells in the testis. The sequence similarity between PH-30 α , PH-30 β , HR1B, and disintegrins indicates a common ancestry. PH-30 α and β are unique, however, in that they contain transmembrane anchors. They may thus be the first recognized members of a novel family of membrane-bound integrin ligands.

The PH-30 α/β complex (Fig. 4c) seems to be the first recognized cellular analogue of a viral fusion protein. In addition to their relevance to understanding membrane fusion, cell-cell adhesion and signal transduction, our findings suggest a new contraceptive strategy. Snake venom disintegrins and peptides derived from them are possible therapeutic anticoagulants^{23,25}. By analogy, peptides corresponding to the PH-30 β disintegrin domain may be useful as contraceptives. □

Received 6 November 1991; accepted 28 January 1992.

1. Snell, W. J. *Curr. Opin. Cell Biol.* **2**, 821-832 (1990).
2. White, J. M. & Blobel, C. P. *Curr. Opin. Cell Biol.* **1**, 934-939 (1989).
3. Stegmann, T., Doms, R. W. & Helenius, A. *Rev. Biophys. biophys. Chem.* **18**, 187-211 (1989).
4. White, J. M. *A. Rev. Physiol.* **52**, 675-697 (1990).

5. Primakoff, P., Hyatt, H. & Tredick-Kline, J. *J. Cell Biol.* **104**, 141–149 (1987).
6. Blobel, C. P., Myles, D. G., Primakoff, P. & White, J. M. *J. Cell Biol.* **111**, 69–78 (1990).
7. Frohman, M. A. & Martin, G. R. *Technique* **1**, 165–170 (1989).
8. Lathrop, W. F., Carmichael, E. P., Myles, D. G. & Primakoff, P. *J. Cell Biol.* **111**, 2939–2949 (1990).
9. Harter, C., James, P., Bächli, T., Semenza, G. & Brunner, J. *J. Biol. Chem.* **264**, 6459–6464 (1989).
10. Barlow, D. J. & Thornton, J. M. *J. molec. Biol.* **201**, 601–619 (1988).
11. Henderson, R. et al. *J. molec. Biol.* **213**, 899–929 (1990).
12. Katow, S. & Sugiyama, A. *J. gen. Virol.* **69**, 2797–2807 (1988).
13. Vidgren, G., Takkinen, K., Kalkkinen, N., Kääriäinen, L. & Pettersson, R. *J. gen. Virol.* **68**, 2347–2357 (1987).
14. Sander, C. & Schneider, R. *Proteins* **9**, 56–68 (1991).
15. Shebuski, R. J., Ramjit, D. R., Bencen, G. H. & Polokoff, M. A. *J. Biol. Chem.* **264**, 21550–21556 (1989).
16. Scarborough, R. M. et al. *J. Biol. Chem.* **266**, 9359–9362 (1991).
17. Dennis, M. S. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2471–2475 (1989).
18. Gan, Z. R., Gould, R. J., Friedman, P. A. & Polokoff, M. A. *J. Biol. Chem.* **263**, 198–19832 (1988).
19. Phillips, D. R., Charo, I. F. & Scarborough, R. M. *Cell* **65**, 359–362 (1991).
20. Hynes, R. O. *Cell* **48**, 549–554 (1987).
21. Ruoslahti, E. & Pierschbacher, M. D. *Science* **238**, 491–497 (1987).
22. Garsky, V. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4022–4026 (1989).
23. D'Souza, S. E., Ginsberg, M. H. & Plow, E. F. *Trends biochem. Sci.* **16**, 246–250 (1991).
24. Saudek, V., Atkinson, R. A. & Pelton, J. E. *Biochemistry* **30**, 7369–7372 (1991).
25. Adler, M., Lazarus, R. A., Dennis, M. S. & Wagner, G. *Science* **253**, 445–448 (1991).
26. Takeya, H., Oda, K., Miyata, T., Omori-Sato, T. & Iwanaga, S. *J. Biol. Chem.* **265**, 16068–16073 (1990).
27. Calarco, P. *J. Electron Microscope Tech.* **17**, 401–411 (1991).
28. Boldt, J., Gunter, L. E. & Howe, A. M. *Gamete Res.* **23**, 91–101 (1989).
29. Bronson, R. A. & Fusi, F. *Biol. Reprod.* **43**, 1019–1025 (1990).
30. Lam, C. T. et al. *J. Biol. Chem.* **262**, 947–950 (1987).
31. D'Souza, S. E., Ginsberg, M. H., Burke, T. A. & Plow, E. F. *J. Biol. Chem.* **265**, 3440–3446 (1990).
32. Wasserman, P. M. *Nature* **235**, 553–560 (1987).
33. Yanagimachi, Y. *Curr. Top. Membr. Transp.* **32**, 3–43 (1988).
34. Miyazaki, S. *J. Reprod. Fert. Suppl.* **42**, 163–175 (1990).
35. Werb, Z., Tremble, P. M., Behrendtsen, O., Crowley, E. & Damsky, C. H. *J. Cell Biol.* **109**, 877–889 (1989).
36. Davis, C. G. *New Biologist* **2**, 410–419 (1990).
37. Matsudaira, P. *J. Biol. Chem.* **262**, 10035–10038 (1987).
38. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995 (1984).
39. Pearson, W. *Meth. Enzym.* **183**, 63–98 (1990).
40. Marck, C. *Nucleic Acids Res.* **16**, 1829–1836 (1988).

ACKNOWLEDGEMENTS. We thank F. Bazan, J. Finer-Moore, P. Garcia, J. Gilbert, J. Glenn, B. Gumbiner, Z. Hall, R. Kelly, G. Kemble, P. McCreary, R. Myers, E. Napolitano, R. Stroud and L. Taylor for suggestions. C. Damsky, R. Fletterick, L. Reichardt and P. Walter for critical reading of the manuscript and H. Czerwinka for help with its preparation. This work was supported by the Muscular Dystrophy Association and the Pew Memorial Trust (to J.M.W.) and NIH (to D.G.M.). T.G.W. is a predoctoral fellow of NSF. The nucleotide sequence of mature PH-30 α and β will appear in the EMBL Sequence Database under the accession numbers Z11719 and Z11720 respectively.

P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips

Meta J. Kuehn*, John Heuser†, Staffan Normark* & Scott J. Hultgren*‡

* Department of Molecular Microbiology, Box 8230 and † Department of Cell Biology and Physiology, Box 8228, Washington University School of Medicine, St. Louis, Missouri 63110, USA

ESCHERICHIA coli is a frequent cause of several common bacterial infections in humans and animals, including urinary tract infections, bacteraemia and bacteria-related diarrhoea and is also the main cause of neonatal meningitis¹. Microbial attachment to surfaces is a key event in colonization and infection and results mainly from a stereochemical fit between microbial adhesins and complementary receptors on host cells. Bacterial adhesins required for extracellular colonization by Gram-negative bacteria are often minor components of heteropolymeric fibres called pili which must be oriented in an accessible manner in these structures to be able to bind to specific receptor architectures. P pili mediate the binding of uropathogenic *E. coli* to a digalactoside receptor determinant present in the urinary tract epithelium. We report here that the adhesin is a component of distinct fibrillar structures present at the tips of the pili. These virulence-associated tip fibrillae are thin, flexible polymers composed mostly of repeating subunits of PapE that frequently terminate with the α -D-galactopyranosyl-(1-4)- β -D-galactopyranose or Gal α (1-4)Gal binding PapG adhesin.

The biosynthesis and expression of P pili involves 11 genes

organized in the *pap* gene cluster on the chromosome in clinical isolates expressing P pili^{1,2}. The DNA sequence of the entire *pap* gene cluster has been determined and the general function of many gene products have been described (Fig. 1). Repeating PapA subunits, probably arranged in a right-handed helical rod, compose the bulk of the pilus³. The PapG adhesin, PapE and PapF are encoded by genes at the distal end of the operon and previously have been localized at the tips of pili⁴. Pilus assembly depends on PapD, a periplasmic chaperone which modulates the productive interactions of the subunits during pilus biogenesis^{5–7}, and PapC, which probably functions as an outer-membrane assembly platform⁸.

High-resolution electron microscopy of the uropathogenic *E. coli* strain J96 (ref. 1) and of *E. coli* HB101 carrying plasmid pPAP5 (which contains the entire *pap* gene cluster) showed that P pili are composite structures consisting of two different fibres joined end to end (Fig. 2). The rigid pilus stalk appeared to be arranged in a helix roughly 10 nm in diameter. A thin fibrillum, roughly 3 nm in diameter, extended from the distal tip of the pilus stalk (Fig. 2, see sections labelled P pili and WT). A single pilus fragment never had more than one tip fibrillum. The lengths of the fibrillae were variable, averaging 42 ± 16 nm for 100 fibrillae examined. Twists occurring every 15 nm indicated that tip fibrillae are open helical structures. The many bends and flexures displayed by the fibrillae in electron micrographs suggested that tip fibrillae are also flexible structures.

The composition of the tip fibrillum was deduced by an electron microscopic investigation. The architecture of purified

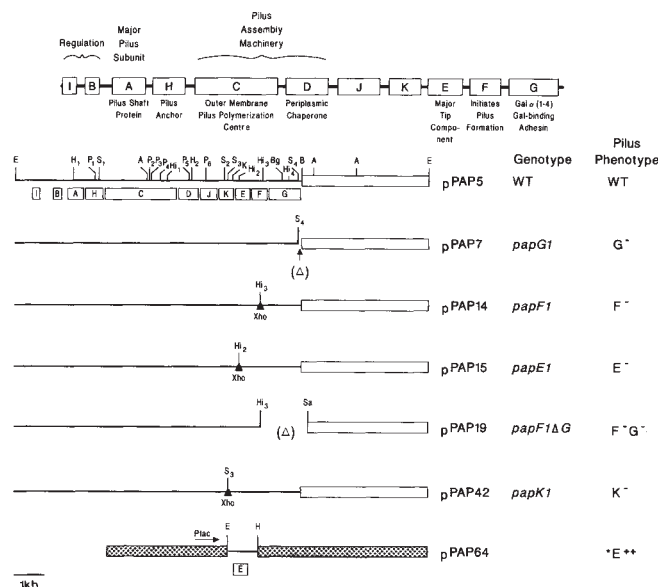


FIG. 1 Summary of the regulation and structure of the 11 genes in the *pap* operon. The established and postulated functions of the gene products are indicated and discussed in the text (for review, see ref. 1). The plasmid pPAP5 (ref. 19) contains the *pap* gene cluster in the vector pBR322 (open bar) and confers expression of P pili to the *E. coli* strain HB101. pPAP7 has a deletion of 42 base pairs at the 3' end of *papG* (Δ)⁹. pPAP14 (ref. 9), pPAP15 (ref. 9) and pPAP42 (ref. 10) contain *Xho* linker insertions (Δ) at the restriction sites indicated, creating the *papF1*, *papE1* and *papK1* genetic lesions, respectively. pPAP19 was constructed by deleting the *Xho*–*Sal*I fragment from pPAP14 and religating⁹. pPAP64 was constructed by amplifying *papE* (from codon 7,402 to 8,109) by polymerase chain reaction (PCR) from pPAP5 as a 0.6-kilobase (kb) *Eco*R1, *Bam*H1 fragment and cloned downstream of the *tac* promoter in the vector pMMB91 (shaded bar)^{5,20}. Restriction endonuclease sites are as follows: A, *Acc*I; B, *Bam*H1; Bg, *Bgl*II; E, *Eco*R1; H, *Hind*III; Hi, *Hinc*II; K, *Kpn*I; P, *Pst*I; S, *Sma*I; Sa, *Sal*I; subscripts number the sites in each group. * E^{++} plus phenotype resulted from complementation of pPAP64 with pPAP5, pPAP15 or pPAP19.

‡ To whom correspondence should be addressed.

wild-type pili was compared with mutant pili missing either PapG (G^- pili), PapF (F^- pili) or PapE (E^- pili). Pili used for this study were purified from individual *E. coli* HB101 strains containing plasmids pPAP5, pPAP7 (*papG1*), pPAP14 (*papF1*) or pPAP15 (*papE1*). These plasmids contained the entire *pap* gene cluster with the latter three having nonpolar *XhoI* linker

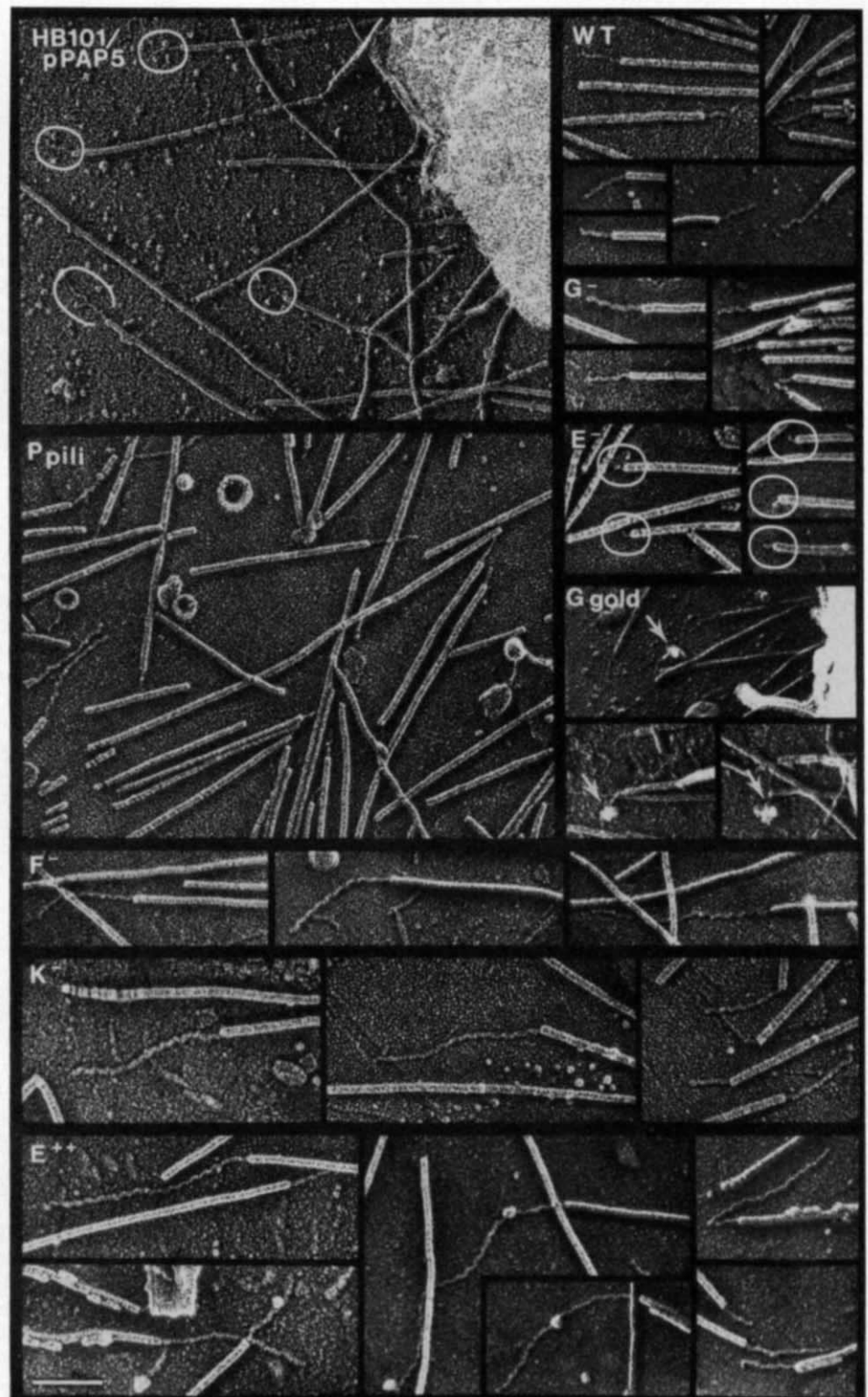
insertions in the indicated genes⁹ (Fig. 1). In addition, pili were purified from HB101/pPAP42 (*papK1*), which has an *XhoI* linker insertion in a new gene designated *papK* (ref. 10) (K^- pili).

Electron microscopy of E^- pili revealed that they lacked tip fibrillae (Fig. 2, section E^-). Some of the E^- pilus rods terminated in short fibrillar stubs, roughly 10-nm long. A normal length

FIG. 2 Freeze-etch electron microscopy revealed that P pili are composite fibres. Thick black lines delineate groups of electron micrographs of piliated bacteria or purified pili of the indicated pilus phenotype. The thinner lines separate each phenotypic section into different microscopic fields of the same pilus phenotype. Pilus phenotypes are identified in the upper left hand corner of each section. The two largest sections in the upper left hand corner are electron micrographs of piliated bacteria (HB101/pPAP5) and purified pili from those bacteria (P pili). Tip fibrillae on pili produced by HB101/pPAP5 are circled. Purified pili from HB101/pPAP5 are shown in the section labelled WT. The sections labelled G^- , E^- , F^- and K^- display electron-micrograph fields of purified pili from HB101 strains carrying pPAP7 (*papG1*), pPAP15 (*papE1*), pPAP14 (*papF1*) and pPAP42 (*papK1*), respectively. E^- pili lack tip fibrillae but terminate in short fibrillar stubs, circled in the E^- section. The micrographs displayed in the E^{++} section are representative of pili purified from HB101 strains where PapE was overexpressed on pPAP64 in the presence of either pPAP5 (WT), pPAP15 (*papE1*) or pPAP19 (*papF1ΔG*). The G-gold section shows the localization of PapG to the tip fibrillae of pili produced by HB101/pPAP5 using immunogold electron microscopy with anti-PapG antibodies. Gold particles appear as solid, white circles in the electron micrographs (arrows) and were often found associated with the distal ends of the fibrillae. B, Bacterial cell. Scale bar, 0.1 μ m, applies to all sections; minor size variation is due to variable platinum deposition.

METHODS. HB101/pPAP5 bacteria were grown on CFA agar and prepared for transmission electron microscopy by adsorption to glass coverslips which were then quick-frozen with a home-made liquid helium-cooled 'cryopress', then freeze dried, and rotary replicated with platinum²¹. P pili were purified according to the method described in Fig. 3 and then prepared for electron microscopy by adsorption to mica chips which were quick-frozen, then freeze-fractured and deep etched, rather than completely freeze-dried, before rotary replication with platinum²². Anti-PapG-PapD rabbit antiserum⁷ was absorbed with HB101/pPAP7 (*papG1*) cells according to ref. 23, and IgG was purified²⁴. Immunogold labelling was done by resuspending HB101/pPAP5 grown on CFA agar in 200 μ l PBS (pH 7.4) and then incubating with 10 μ l of the anti-PapD-PapG IgG with rocking for 15 min at 25 °C. Cells were pelleted and washed six times in PBS. Anti-rabbit IgG (5 μ l) coupled to 10 nm gold (Sigma) was added, and

the mixture was incubated for 5 min at 25 °C. Gold-labelled cells were then washed six times with PBS using 0.45 μ m Ultrafree-MC filter unit tubes (Millipore).



tip fibrillum was never detected among the hundreds of E^- pili examined. By contrast, examination of G^- , F^- and K^- pili by electron microscopy showed that the tip fibrillum structure was intact (Fig. 2, sections G^- , F^- and K^-). Thus, PapE was the only gene product essential in the formation of tip fibrillae, suggesting that it was the major component of the tip fibrillum. Tip fibrillae present on F^- and K^- pili appeared to be longer (145 ± 65 nm and 80 ± 27 nm, respectively) than those on wild-type P pili suggesting their possible role in length control. These results could not be explained on the basis of polarity because other studies have shown that these same linker mutant derivatives of pPAP5 were nonpolar on the expression of distal genes⁹.

The hypothesis that repeating subunits of PapE formed the majority of the tip fibrillum was tested by *trans* complementation. The *papE1* lesion carried on the pPAP15 plasmid was complemented in *trans* with the pPAP64 plasmid which carried only the *papE* gene downstream of the inducible *tac* promoter (Fig. 1). Induction of the *tac* promoter with 0.01 mM IPTG (isopropylthiogalactoside) led to the restoration of tip fibrillae as visualized in electron micrographs of the purified pili (Fig. 2, section E^{++}). The strong expression of *papE* from the *tac* promoter resulted in the production of tip fibrillae that were about three times longer than wild type; a sample size of 100 fibrillae had an average length of 113 ± 41 nm. PapF and PapG, encoded by genes downstream of *papE*, were not required for

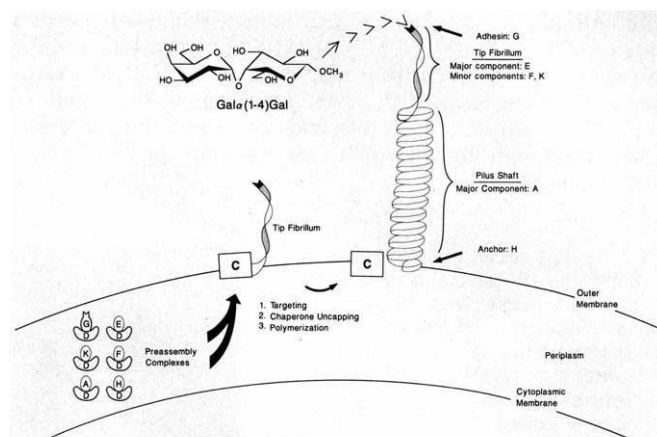


FIG. 4 Model of P pilus biogenesis. Details of the model are discussed in the text. Briefly, the PapD chaperone (D)^{27,28} interacts with each pilus protein subunit forming assembly competent complexes^{1,5-7} which are targeted to the PapC outer membrane assembly protein required for P pilus biogenesis. PapA subunits pack into a right-handed helical rod. PapH incorporation terminates polymerization of PapA and anchors the pilus to the cell²⁶. PapE subunits form open helical fibres called tip fibrillae. The PapG adhesin, positioned at the distal end of the fibrillum (shown in black), mediates binding to the Gal α (1-4)Gal receptor determinant. PapF⁴ and PapK (F. Jacob, J. Tennent, personal communication) have been localized to the tips of pili; their position in the tip fibrillum are not known.

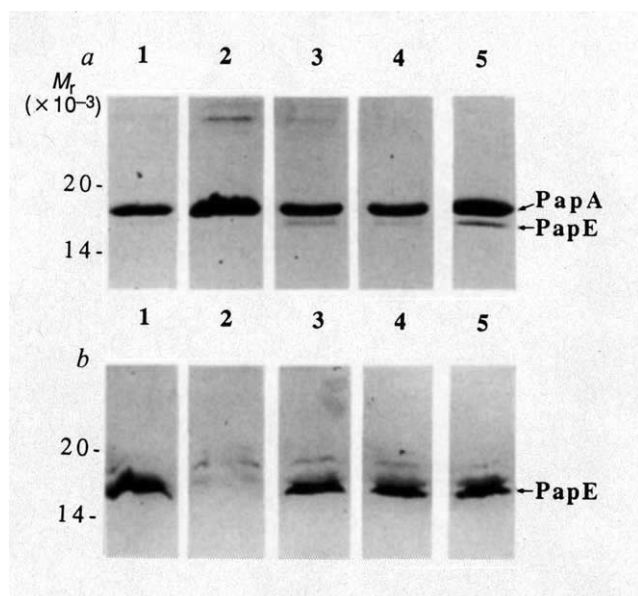


FIG. 3 *Trans* complementation with *papE* converts wild-type P pili, E^- pili and F^- pili into E^{++} pili as determined by Coomassie blue stained SDS-PAGE (a) and immunoblotting using anti-PapE antisera (b). Lanes 1, 1.5 μ g wild-type P pili purified from HB101/pPAP5; 2, 3 μ g E^- pili purified from HB101/pPAP15 (*papE1*); 3, 1.5 μ g E^{++} pili purified from HB101/pPAP5+pPAP64; 4, 1.5 μ g E^{++} pili purified from HB101/pPAP15 (*papE1*)+pPAP64; 5, 2 μ g E^{++} pili purified from HB101/pPAP19 (*papF1ΔG*)+pPAP64. Amounts analysed are approximate owing to low reactivity of pili in protein determination assays. PapA, PapE and molecular mass (M_r) standards are indicated.

METHODS. *E. coli* harbouring *pap*-containing plasmids were grown on CFA agar containing 100 μ g ml^{-1} carbenicillin and passaged overnight at 37 °C at least two successive times. HB101/pPAP5+pPAP64, HB101/pPAP15+pPAP64 and HB101/pPAP19+pPAP64 were grown on CFA agar containing 0.01 mM IPTG, 100 μ g ml^{-1} carbenicillin and 50 μ g ml^{-1} kanamycin. Bacteria were resuspended in 0.5 mM Tris and 75 mM NaCl, pH 7.2. Pili were heat extracted²⁵ and then precipitated with 150 mM NaCl and 200 mM KCl for 2 h at 25 °C, stirring constantly. Precipitated pili were then pelleted and resuspended in 5 mM Tris, pH 8.0. Samples were boiled for 5 min in SDS-PAGE sample buffer before electrophoresis. Quantitation of Coomassie blue stained bands was done by densitometry at 633 nm (LKB Ultrascan XL laser densitometer, Pharmacia).

the formation of fibrillae. *PapF* and *papG* were inactivated and deleted from the *pap* operon, respectively. When the corresponding plasmid (pPAP19; *papF1ΔG*) (see Fig. 1) was complemented with pPAP64, the same long fibrillar phenotype shown in the E^{++} section of Fig. 2 was produced. In addition, overexpressing *papG* from the *tac* promoter on pPAP55 (ref. 5) in the presence of the *pap* operon on pPAP5 had no detectable effect on the fibrillar architecture (data not shown). The complementation data thus strongly support the theory that PapE is the major component of the tip fibrillum and demonstrate its ability to polymerize into a fibre.

PapE subunits expressed in *trans* in the above complementation of the *papE1* lesion presumably restored the fibrillar phenotype by assembling into tip fibrillae. An increased number of PapE subunits probably polymerized into the fibrillae, thus increasing their length. These hypotheses were tested by investigating whether the tip fibrillar phenotype, as determined by electron microscopy, correlated with the presence of PapE in purified pili, as determined in biochemical analyses. E^{++} pili were purified from strains where PapE was overexpressed in the presence of (1) the *pap* operon (HB101/pPAP5+pPAP64); (2) the *papE1* operon (HB101/pPAP15+pPAP64); and (3) the *papF1ΔG* operon (HB101/pPAP19+pPAP64). About 1.5 μ g of wild-type and E^{++} pili, and 3 μ g of E^- pili were analysed by SDS-PAGE and by immunoblotting with anti-PapE antiserum. A PapE band of relative molecular mass 16,500 (M_r 16.5 K) (Fig. 3a) that reacted with the anti-PapE antisera (Fig. 3b) was detected in wild type as well as all three forms of E^{++} pilus preparations. The 16.5K PapE band was absent in E^- pili (Fig. 3a, b, lane 2). The ratio of PapE/PapA was 2.5-fold greater in pilus preparations from HB101/pPAP5+pPAP64 than from HB101/pPAP5 as determined by densitometric scanning of the Coomassie blue stained bands. These data strongly support the conclusion that PapE polymerizes into fibres called tip fibrillae and that the overproduction of PapE results in an increased length of the PapE fibrillar fibres.

The Gal α (1-4)Gal-binding PapG adhesin has been localized previously to the tips of pili⁴, and thus we sought to determine its location in the tip fibrillum. Anti-PapG antibodies, obtained

from rabbits immunized with purified PapG-PapD complexes^{5,7}, were used in association with gold-conjugated anti-rabbit antibodies in an immunogold electron microscopy procedure to localize PapG in pili produced by HB101/pPAP5. Gold particles were found associated with the tip fibrillae and often appeared to be at the distal ends of the fibrillae (see Fig. 2, section G-gold). Anti-PapD antibodies did not bind to the pili, as has been shown previously⁶, arguing that anti-PapG antibodies recognized PapG in the fibrillae. On the basis of these data, we propose that PapG is located at the distal ends of tip fibrillae (see model, Fig. 4).

Filamentous adhesins on bacteria generally have one of two basic morphologies: rod-like pili with a diameter of roughly 7 nm, or flexible, thin fibres with a diameter of 2–5 nm¹¹. X-ray crystallography of the type 1 pilus revealed that it was composed of subunits arranged in a rigid, right-handed helix with 3 $\frac{1}{8}$ subunits per turn and an axial hole of roughly 2 nm (ref. 12). Other kinds of rod-like pili, such as P, S, 987P, CFA1, CS1, CS2 and F7 pili of *E. coli*, probably consist of subunits packed into a three-dimensional array similar to the type 1 pilus. By contrast, K88, K99, F41 and CS3 fimbriae seem to consist of subunits packed into an open helical structure without an axial hole^{11,13}. Each of the above subunit types is related in primary sequence. Thus, differences in the packing of subunits into three-dimensional supramolecular structures probably determines the respective architectural differences among these structures. Generally, adhesins seem to be components of loosely coiled or linear polymers as is the case for K88 and K99 pili^{14,15}, and P pilus tip fibrillae.

Located in the flexible tip fibrillum, the PapG adhesin probably has maximal steric freedom in recognizing glycolipid receptors on eukaryotic cells (see model, Fig. 4). Similarly, Protein 38, the receptor-binding protein from bacteriophages T2, K3 and OX2 (ref. 16), is located at the tapered, distal ends of the long tail fibre¹⁷. Our studies suggest that P pili may have evolved from two genetic sources, one encoding long, rigid shafts, and another encoding adhesive fibrillar polymers. The introduction of adaptor proteins may have united the fibrillae to the stalk structures. The ability of PapE to bind to immobilized fibronectin¹⁸ suggests that tip fibrillae have evolved into multifunctional virulence determinants that were joined to the ends of P pili to increase their accessibility to host receptors. □

Received 25 October; accepted 24 December 1991.

- Hultgren, S. J., Abraham, S. N. & Normark, S. A. *Rev. Microbiol.* **45**, 383–415 (1991).
- Hultgren, S. J. & Normark, S. *Curr. Opin. Genet. Dev.* **1**, 313–318 (1991).
- Baga, M. *et al. J. Bact.* **157**, 330–333 (1984).
- Lindberg, F., Lund, B., Johansson, L. & Normark, S. *Nature* **328**, 84–87 (1987).
- Hultgren, S. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 4357–4361 (1989).
- Lindberg, F., Tennent, J. M., Hultgren, S. J., Lund, B. & Normark, S. *J. Bact.* **171**, 6052–6058 (1989).
- Kuehn, M. J., Normark, S. & Hultgren, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10586–10590 (1991).
- Norgren, M., Baga, M., Tennent, J. M. & Normark, S. *Molec. Microbiol.* **1**, 169–178 (1987).
- Lindberg, F., Lund, B. & Normark, S. *EMBO J.* **3**, 1167–1173 (1984).
- Lindberg, F. thesis, Umeå Univ. (1987).
- De Graaf, F. K. & Mooi, F. R. *Adv. Microb. Physiol.* **28**, 65–143 (1986).
- Brinton, C. C. *Trans. N.Y. Acad. Sci.* **27**, 1003–1054 (1965).
- De Graaf, F. K., Klemm, P. & Gaastra, W. *Infect. Immun.* **33**, 877–883 (1981).
- Roosendaal, B., Gaastra, W. & De Graaf, F. K. *FEMS Microbiol. Lett.* **22**, 253–258 (1984).
- Jacobs, A. A. C., Venema, J., Leeven, J., Van Pelt-Heerschap, H. & De Graaf, F. K. *J. Bact.* **169**, 735–741 (1987).
- Riede, I., Drexler, K., Schwarz, H. & Henning, U. *J. molec. Biol.* **194**, 23–30 (1987).
- Ward, S. *et al. J. molec. Biol.* **54**, 15–31 (1970).
- Westerlund, B. *et al. Molec. Microbiol.* **5**, 2965–2975 (1991).
- Hull, R. A., Gill, R. E., Hsu, P., Minshaw, B. H. & Falkow, S. *Infect. Immun.* **33**, 933–938 (1981).
- Furste, J. P. *et al. Gene* **48**, 119–131 (1986).
- Heuser, J. *J. Cell Biol.* **108**, 401–411 (1989).
- Heuser, J. *J. Electron Microsc. Tech.* **13**, 244–263 (1989).
- Lindberg, F., Lund, B. & Normark, S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1891–1895 (1986).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual*, 310 (Cold Spring Harbor Laboratory Press, New York, 1988).
- Hoschutsky, H., Lottspeich, F. & Jann, K. *Infect. Immun.* **57**, 76–81 (1989).
- Baga, M., Norgren, M. & Normark, S. *Cell* **49**, 241–251 (1987).
- Holmgren, A. & Branden, C.-I. *Nature* **342**, 248–256 (1989).
- Holmgren, A., Kuehn, M. J., Branden, C.-I. & Hultgren, S. *EMBO J.* (in the press).

ACKNOWLEDGEMENTS. We thank J. Pinkner and M. J. Wick for the construction of pPAP64, R. Roth for technical assistance, K. Dodson and S. Abraham for discussions and F. Jacob and J. Tennent for the information that PapK is a pilus tip protein. This work was supported by the National Institute of Health (USPHS) and the Lucille P. Markey Charitable Trust.

Cardioviral internal ribosomal entry site is functional in a genetically engineered dicistronic poliovirus

Akhteruzzaman Molla, Sung Key Jang*, Aniko V. Paul, Quentin Reuer* & Eckard Wimmer

Department of Microbiology, School of Medicine, Health Sciences Center, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

HIGH mutation rates have driven RNA viruses to shorten their genomes to the minimum possible size¹. Mammalian (+)-strand RNA viruses and retroviruses have responded by reducing the number of *cis*-acting regulatory elements, a constraint that has led to the emergence of the polyprotein². Poliovirus is a (+)-stranded picornavirus whose polyprotein, encoded by an open reading frame spanning most of the viral RNA³, is processed by virus-encoded proteinases^{4,5}. Despite their genetic austerity, picornaviruses have retained long 5' untranslated regions^{6–8}, which harbour *cis*-acting elements that promote initiation of translation independently of the uncapped 5' end of the viral messenger RNA^{9–12}. These elements are termed 'internal ribosomal entry sites'¹⁰ and are formed from highly structured RNA segments^{13–15} of at least 400 nucleotides¹⁶. How these elements function is not known, but special RNA-binding proteins may be involved¹⁷. The ribosome or its 40S subunit probably binds at or near a Y_nX_mAUG motif (where Y is a pyrimidine and X is a purine) at the 3' border of the internal ribosomal entry site¹⁷, which either provides the initiating codon^{16,18} or enables the ribosome to translocate to one downstream (E.W. *et al.*, submitted). Initiation from most eukaryotic messenger RNAs usually occurs by ribosomal recognition of the 5' and subsequent scanning to the AUG codon¹⁹. Here we describe a genetic strategy for the dissection of polyproteins which proves that an internal ribosomal entry site element can initiate translation independently of the 5' end.

We have interrupted the open reading frame (ORF) of the poliovirus type 1 (Mahoney) (PV1(M)) polyprotein by inserting 573 nucleotides of the encephalomyocarditis virus (EMCV) 5' nontranslated region (NTR) (nucleotides 260–833) at the P1/P2 junction which is normally cleaved by viral proteinase 2A^{pro} (ref. 20) (Fig. 1a, b). The insert contained the EMCV internal ribosomal entry site (IRES) (ref. 16), 15 nucleotides of the EMCV ORF, and an added codon for tyrosine, the latter to provide a *cis* cleavage site for 2A^{pro} (Fig. 1d). The ORF of capsid precursor P1 was terminated by a UGA triplet²¹. Altogether, 602 nucleotides (of non-coding RNA) were placed between domains P1 and P2 (Fig. 1b) and the resulting construct, containing all of the poliovirus complementary DNA, was incorporated into a phage T7 transcription vector²².

The ability of the dicistronic viral RNA to direct synthesis of two polyproteins under the control of two different IRES elements was tested by translation of pT7PVE2A transcripts in a HeLa cell-free extract²³. The pattern of translation and the processing of products (Fig. 1e, lane 3) resembled that directed by transcripts of pT7PVXL2 (lane 2), an observation suggesting that the IRES elements of PV1 and EMCV function with comparable efficiency in a HeLa cell extract⁹. The band of polypeptide 2A^{pro} produced by the dicistronic RNA (referred to as 2A', lane 3) migrated more slowly than wild-type 2A^{pro}, which we attribute to the five amino acids originating from the EMCV polyprotein plus the Tyr residue (Fig. 1d). Apparently the N-terminal extension of 2A' was not removed. The presence of

* Present addresses: Department of Life Sciences, Pohang Institute of Science and Technology, PO Box 125, Pohang 790-600, Korea (S.K.J.); Biology Department, University of Alaska, 3211 Providence Drive, Anchorage, Alaska 99508, USA (Q.R.).

a few amino acids at the N terminus of 2A^{pro} is not detrimental to proteinase activity²⁴, although it seems that 2A' cannot produce 3D' and 3C', the products of a *trans* cleavage of 3CD (ref. 5) (3C' does not resolve in Fig. 1e), which we cannot explain.

Transfection of HeLa cell monolayers with transcripts of pT7PVXL2 or pT7PVE2A produced characteristic cytopathic effects 18 and 36 h later, respectively. Poliovirus W1-P1/E/P2,3-1, isolated from cells transfected with pT7PVE2A RNA, exhibited a small-plaque phenotype (Fig. 2). Its RNA was analysed by cDNA synthesis and amplification by polymerase chain reaction (PCR)²³ to yield a fragment of the expected length of 780 nucleotides (178 nucleotides from the polio genome sequence plus 602 nucleotides from the insert). The genotype

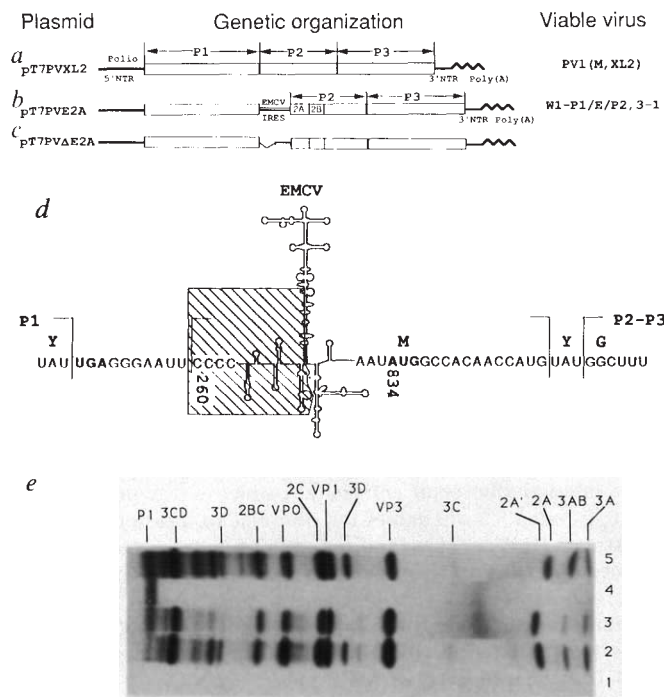


FIG. 1 Genetic organization of poliovirus and its dicistronic variants. Shown is the genetic content of the hypothetical viruses (open bars, reading frames; P1, capsid precursor; P2 and P3, non-structural proteins⁶) that were engineered into plasmids containing the promoter for phage T7 RNA polymerase²², and viruses recovered after transfection of HeLa cell monolayers with transcripts spanning the poliovirus cDNAs (minus sign, no virus recovered). *a*, Parental construct consisting of pT7PV1-5 (ref. 22), of which the 3' fragment (*Bgl*III site at nucleotide 5,600, to *Eco*RI, immediately downstream of the poly(A)) had been exchanged with a corresponding fragment from plasmid T7pGEM polio (supplied by P. Sarnow). *b*, Dicistronic constructs in which the polyprotein was interrupted by 602 nucleotides of non-coding RNA containing the EMCV IRES. 2A and 2B are gene products mapping to the N-terminal portion of P2.2A is the proteinase normally cleaving P1/P2. *c*, As *b*, except that a segment of the EMCV IRES was deleted (nucleotide 260 to 434) as shown in *d* (shaded area). *d*, Detail of the insert between domains P1 and P2 of the polyprotein. Capital letters Y, M and G (bold) are single-letter abbreviations for amino acids. Numbers refer to the boundaries of the segment of the 5'NTR of EMCV RNA¹⁶. Line drawing depicts the secondary structure of the EMCV IRES (ref. 17). Termination of the P1 ORF is described in ref. 21. *e*, Polyacrylamide gel (12.5%) electrophoresis of poliovirus-specific proteins synthesized under the direction of T7 transcripts in a HeLa cell-free extract. Lane 1, no RNA; lane 2, pT7PVXL2 RNA; lane 3, pT7PVE2A RNA; lane 4, pT7PVE2A RNA; lane 5, poliovirus proteins synthesized in HeLa cells after infection with (PV1(M, XL2) as markers. 2A, Wild-type 2A^{pro}; 2A', 2A^{pro} with additional amino acids at the N terminus originating from the ORF of the insert (see *d*).

METHODS. Construction of the plasmids will be described elsewhere (A.M. *et al.*, manuscript in preparation). Plasmids were transcribed and the RNAs translated *in vitro*²³. Optimal conditions for *in vitro* translation of poliovirus and EMCV RNAs are very different (ref. 9, and refs therein)—conditions used here are suboptimal for the function of the IRES element of both these viruses.

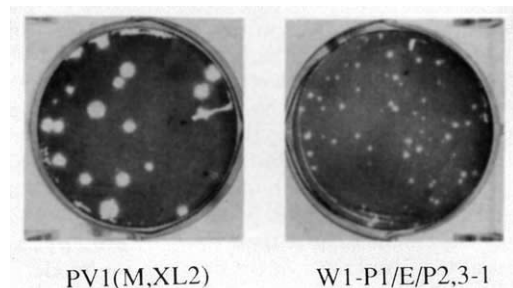


FIG. 2 Plaque morphology of PV1(M, XL2) and W1-P1/E/P2,3-1 viruses. The reason for the small-plaque phenotype of the dicistronic poliovirus is unknown but may be the result of delayed RNA replication, A.M., S.K.J., A.P., Q.R. and E.W., manuscript in preparation. Another factor may be encapsidation as the genome of this variant is 8% larger than that of wild-type poliovirus. For methods, see ref. 29.

was confirmed by partial sequencing of viral RNA and PCR-amplified cDNA (data not shown).

Comparison of protein synthesis of parental and dicistronic viruses in HeLa cells (Fig. 3) indicated that the time course of the appearance of viral polypeptides and the shutoff of host cell protein synthesis was similar, although protein synthesis by PV1(M, XL2) was slightly earlier (compare lanes 9 with 10). As already noted in cell-free translation, the dicistronic virus produced a 2A^{pro} (designated 2A') larger than the 2A^{pro} of the wild-type virus, but *in vivo* it produced the *trans* cleavage products of 3CD by 2A^{pro}.

Perhaps translation of the second cistron in W1-P1/E/P2,3-1 was by a termination/reinitiation mechanism¹⁹, in which translation is terminated at the end of the P1 cistron, followed by scanning of 573 nucleotides until the initiation codon of the

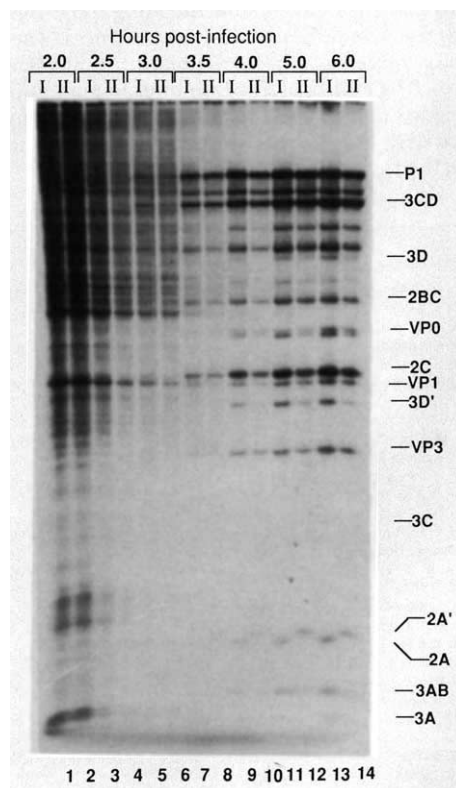
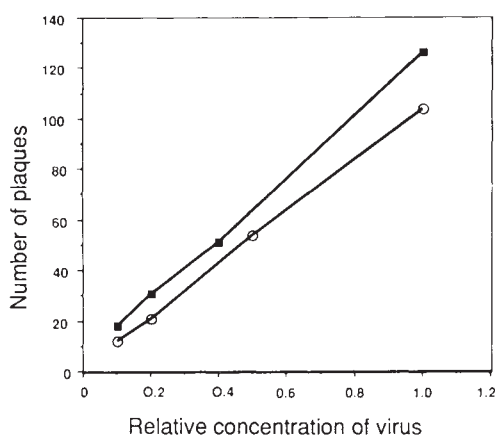


FIG. 3 Time course of appearance of poliovirus-specific proteins after infection with PV1(M, XL2) (I) and W1-P1/E/P2,3-1 (II). **METHODS.** HeLa R-19 monolayer cells were infected with 5 PFU per cell of either virus. At the indicated times post-infection, the cells were incubated for 10 min with [³⁵S]Met (300 uCi ml⁻¹, 1,000 Ci mmol⁻¹), and the labelled products were analysed on 12.5% SDS-polyacrylamide gels²³.

second polypeptide is reached. If this were the case, the scanning ribosomal subunit would have to ignore ten intervening noncoding AUG triplets, of which one is in a perfect context for initiation^{16,18}. To investigate this possibility, we constructed a dicistronic viral genome (Fig. 1c) with a deletion (nucleotides 260–434) in the EMCV IRES (cross-hatched in Fig. 1d). Translation of transcripts of this construct *in vitro* gave only unprocessed polypeptide P1 (Fig. 1e, lane 4) and no P2/P3 polypeptides (including the proteinases necessary for P1 processing). Transfection of cells with these transcripts produced no detectable cytopathic effects, and at 36 h after transfection we could detect a DNA fragment of the expected 606 nucleotides (178 from the viral RNA plus 428 from the insert) after cDNA synthesis and PCR amplification of total RNA from the transfected cells (data not shown). The deletion should probably have enhanced rather than blocked a termination/reinitiation mechanism, particularly as several hairpins were removed. We have therefore discounted this mechanism.

The proliferation of dicistronic virus W1-P1/E/P2,3-1 can also be explained as follows. If one cell were infected with many virus particles, some of the viral genomes could be fragmented in the EMCV IRES, thereby facilitating 5' end-dependent translation of the non-structural proteins. Plaque formation could then result from complementation, the cleaved genomes providing viral proteins to replicate uncleaved genomes, while intact genomes synthesize capsid proteins to form progeny. The requirement of genome fragmentation and complementation for dicistronic virus replication can be excluded by measuring the relation between virus concentration and plaque formation. If the relation is linear, each individual plaque must originate from a single poliovirus particle²⁵. We find such a linear relation after a series of titrations with W1-P1/E/P2,3-1 (Fig. 4).



Dilution	Plaque count	Titre (PFU ml ⁻¹)
Experiment 1	18	1.8 × 10 ⁸
	31	1.5 × 10 ⁸
	51	1.3 × 10 ⁸
	126	1.3 × 10 ⁸
Experiment 2	12	6.0 × 10 ⁷
	21	5.3 × 10 ⁷
	54	5.4 × 10 ⁷
	104	5.2 × 10 ⁷

FIG. 4 Proportionality between plaque count and concentration of poliovirus W1-P1/E/P2,3-1. Plaque numbers are the average of those from duplicate wells. Plaque assays were as described²⁹. Experiments 1 (squares) and 2 (circles): dilutions were made from virus stocks obtained after the first and second passage through HeLa cells, respectively.

The proliferation of poliovirus W1-P1/E/P2,3-1, whose polypeptide is interrupted by a non-coding sequence containing an IRES element, is remarkable. This means that a eukaryotic RNA virus with two consecutive ORFs, rather than overlapping reading frames, can exist in principle. However, such an entity has not been found before, presumably because of the genetic pressure to shorten viral RNA genomes¹. On the other hand, synthesis of two distinct polypeptide species by RNA viruses, of which one generates a proteinase cleaving both polypeptides, is common. Examples are all retroviruses and many plus-strand RNA viruses, most notably cow pea mosaic virus (CPMV)⁴. The dicistronic poliovirus and CPMV have similar gene order and amino-acid sequences²⁶, the main difference being that the genome of CPMV is bipartite.

The infectivity of the dicistronic poliovirus W1-P1/E/P2,3-1 relative to the number of particles shows that a single virus particle can initiate an infectious cycle. This, together with the phenotype of the deletion mutant, is evidence that the highly structured segment of the EMCV IRES can initiate translation internally *in vivo*. This mechanism of initiation deviates from that commonly accepted for 5' cap-dependent translation^{17,27}, but in both modes of initiation the translational machinery is attracted to the mRNA upstream of the ORF. It may not be surprising if internal ribosomal entry also functions in the translation of specific cellular mRNA species. This has been demonstrated for the mRNA of immunoglobulin heavy-chain binding protein²⁸, although the structure of the 5'NTR of this cellular mRNA has no apparent resemblance to that of picornaviral mRNAs.

The dissection of a protein using an IRES element to yield separate polypeptides is a novel method for analysing protein function genetically. For viral polypeptides, it can be applied to investigate precursors in the proteolytic processing pathway and their products. In principle, it can be used to scan for independently functioning domains of any polypeptide. □

Received 11 November; accepted 31 December 1991.

- Reaney, D. C. *Symp. Soc. gen. Microbiol.* **35**, 175–196 (1984).
- Kräusslich, H.-G. & Wimmer, E. *A. Rev. Biochem.* **57**, 701–754 (1988).
- Kitamura, N. *et al. Nature* **291**, 547–553 (1981).
- Hellen, C. U. T., Kräusslich, H.-G. & Wimmer, E. *Biochemistry* **28**, 9881–9890 (1989).
- Harris, K. S., Hellen, C. U. T. & Wimmer, E. *Seminars in Virology* Vol. I (ed. Strauss, J. H.) 323–333 (Saunders, Philadelphia, PA, USA, 1990).
- Rueckert, R. R. in *Virology* 2nd edn (eds Fields, B. N. *et al.*) 507–548 (Raven Press, New York, 1990).
- Dorner, A. J., Dorner, L. F., Larsen, G. R., Wimmer, E. & Anderson, C. W. *J. Virol.* **42**, 1017–1028 (1982).
- Palmenberg, A. C. in *Positive Strand RNA Viruses, UCLA Symposia on Molecular and Cellular Biology* Vol. 54 (eds Brinton, M. A. & Rueckert, R. R.) 25–34 (A. R. Liss, New York, 1987).
- Jang, S. K. *et al. J. Virol.* **62**, 2636–2643 (1988).
- Jang, S. K., Davis, M. V., Kaufman, R. J. & Wimmer, E. *J. Virol.* **63**, 1615–1660 (1989).
- Pelletier, J. & Sonenberg, N. *Nature* **334**, 320–325 (1988).
- Pelletier, J. & Sonenberg, N. *J. Virol.* **63**, 441–444 (1989).
- Pilipenko, E. V. *et al. Virology* **168**, 201–209 (1989).
- Pilipenko, E. V., Blinov, V. M., Chernov, B. K., Dmitrieva, T. M. & Agol, V. I. *Nucleic Acids Res.* **17**, 5701–5711 (1989).
- Skinner, M. A. *et al. J. molec. Biol.* **207**, 379–392 (1989).
- Jang, S. K. & Wimmer, E. *Genes Dev.* **4**, 1560–1572 (1990).
- Jang, S. K., Pestova, T., Hellen, C. U. T., Witherell, G. W. & Wimmer, E. *Enzyme* **44**, 292–309 (1990).
- Kaminski, A., Howell, M. T. & Jackson, R. J. *EMBO J.* **9**, 3753–3759 (1990).
- Kozak, M. *J. Cell Biol.* **108**, 229–241 (1989).
- Hellen, C. U. T., Fäcke, M., Kräusslich, H.-G., Lee, C.-K. & Wimmer, E. *J. Virol.* **65**, 4226–4231 (1991).
- Nicklin, M. J. H., Kräusslich, H.-G., Toyoda, H., Dunn, J. J. & Wimmer, E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4002–4006 (1987).
- Van der Werf, S., Bradley, J., Wimmer, E., Studier, F. W. & Dunn, J. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2330–2334 (1986).
- Molla, A., Paul, A. V. & Wimmer, E. *Science* **254**, 1647–1651 (1991).
- Kräusslich, H.-G., Nicklin, M. J. H., Toyoda, H., Etchison, D. & Wimmer, E. *J. Virol.* **61**, 2711–2718 (1987).
- Davis, B. D., Dulbecco, R., Eisen, H. N. & Ginsberg, H. S. in *Microbiology* 3rd edn 874–883 (Harper and Row, New York, 1980).
- Argos, P., Kramer, G., Nicklin, M. J. H. & Wimmer, E. *Nucleic Acids Res.* **12**, 7251–7267 (1984).
- Sonenberg, N. *Curr. Top. Microbiol. Immun.* **161**, 23–47 (1990).
- Macejak, D. C. & Sarnow, P. *Nature* **353**, 90–94 (1991).
- Pincus, S. E., Diamond, D. C., Emmini, E. A. & Wimmer, E. *J. Virol.* **57**, 638–646 (1986).

ACKNOWLEDGEMENTS. We thank C. U. T. Hellen for suggestions and reviewing the manuscript. P. Sarnow for plasmid T7pGEMpolio, J. Mugavero and C. Witherell for technical help. C. Helmke for photography and C. Rafferty for preparing the manuscript. Supported by the National Cancer Institute and National Institute of Allergy and Infectious Diseases of the NIH.

A human recombinant haemoglobin designed for use as a blood substitute

Douglas Looker, Debbie Abbott-Brown, Paul Cozart, Steven Durfee, Stephen Hoffman, Antony J. Mathews, Jeanne Miller-Roehrich, Steven Shoemaker, Stephen Trimble, Giulio Fermi*, Noboru H. Komiya*, Kiyoshi Nagai* & Gary L. Stetler†

Somatogen Inc., 5797 Central Avenue, Boulder, Colorado 80301, USA

* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

THE need to develop a blood substitute is now urgent because of the increasing concern over blood-transmitted viral and bacterial pathogens¹. Cell-free haemoglobin solutions^{2,3} and human haemoglobin synthesized in *Escherichia coli*⁴ and *Saccharomyces cerevisiae*⁵ have been investigated as potential oxygen-carrying substitutes for red blood cells. But these haemoglobins cannot be used as a blood substitute because (1) the oxygen affinity in the absence of 2,3-bisphosphoglycerate is too high to allow unloading of enough oxygen in the tissues⁶, and (2) they dissociate into $\alpha\beta$ dimers⁷ that are cleared rapidly by renal filtration⁸⁻¹⁰, which can result in long-term kidney damage⁷⁻⁹. We have produced a human haemoglobin using an expression vector containing one gene encoding a mutant β -globin with decreased oxygen affinity and one duplicated, tandemly fused α -globin gene. Fusion of the two α -globin subunits increases the half-life of this haemoglobin molecule *in vivo* by preventing its dissociation into $\alpha\beta$ dimers and therefore also eliminates renal toxicity.

In both oxy- and deoxyhaemoglobin the N-terminal residue of one α -subunit and the C-terminal residue of the other α -subunit are only between 2 and 6 Å apart^{11,12}, a distance that could be spanned by one or two amino acids. An extra amino acid can be added to the C-terminal arginine of the α -subunits by trypsin-catalysed reverse hydrolysis without altering the oxygen-binding properties¹³. Furthermore, because the C-terminal Arg and the N-terminal Val of the two α -subunits are bound to one another through a salt bridge in deoxyhaemoglobin, the tandem fusion of the two α -chains should not greatly affect the protein. The effect of expressing a tandem α -globin on the folding and assembly pathway of haemoglobin *in vivo* cannot be predicted, although secreted F_v antibody fragments with their V_H and V_L domains fused by a 15-amino-acid peptide can be assembled in *E. coli*^{14,15}.

We constructed an *E. coli* expression vector (pSGE1.1-E4) that contains two copies of the α -globin gene fused in tandem by a single codon encoding a glycine residue; this created a fusion junction with the sequence Arg(141 α 1)GlyVal(1 α 2). In this construct a single operon encoding a di- α -globin and a β -globin chain is transcribed from a single TAC promoter⁴. To decrease the oxygen affinity of this engineered haemoglobin (Hb) we introduced an additional Asn-108 β → Lys mutation. In its naturally occurring form (Hb Presbyterian) this substitution reduces the oxygen affinity and slightly increases the Bohr effect¹⁶. Figure 1 shows that on induction of the haemoglobin operon with IPTG, *E. coli* cells carrying the plasmid containing the fused α -globin gene produce two polypeptides corresponding to β -globin and fused di- α -globin. Haemoglobin produced by induced cells accumulates to 5–10% of the total cell protein.

To assess the separate effects of the α -globin fusion and the Asn-108 β → Lys mutation, we constructed vectors to express four related proteins: (1) a mutant form of the tetramer that

TABLE 1 Functional properties of rHb and wild-type human Hb

	HbA ₀	rHb1.1	rHb1.0	rHb0.1	rHb0.0
Bohr coefficient	-0.47	-0.32	-0.35	-0.25	-0.27
P ₅₀ (torr)	4.5	17.2	19.8	4.5	7.3
Hill coefficient (n_{max})	2.9	2.35	2.53	2.04	2.59
α -Chain	WT	di- α	WT	di- α	WT
α -Chain (N _t)	Val Leu	Met Leu	Met Leu	Met Leu	Met Leu
β -Chain (β 108)	Asn	Lys	Lys	Asn	Asn
β -Chain (N _t)	Val His	Met His	Met His	Met His	Met His

The Bohr coefficient is expressed as the slope of the line generated by plotting log P₅₀ as a function of increasing pH (PH 6.8–7.8). Also given are the amino acids at the amino termini of the α - and β -chains, at position 108 of the β -chain, and the structure of the α -chain. The various haemoglobin mutants are: rHb0.0, non-fused α -globin chains, no β -globin mutation; rHb1.0, non-fused α -globin chains, Asn-108 β → Lys mutation; rHb0.1, fused α -globin chains, no β -globin mutation; and rHb1.1, fused α -globin chains, with Asn-108 β → Lys mutation. All forms of rHb carry the Val-1 α → Met and Val-1 β → Met mutations. Recombinant and wild-type human haemoglobins were purified from *E. coli* or red blood cells and their oxygen binding properties tested^{4,5}. Oxygen equilibrium binding data were collected at 25 °C in 50 mM HEPES buffer, pH 7.4, containing 100 mM NaCl. WT, wild type; N_t, amino terminus.

contains Val-to-Met changes at the N termini of both α - and β -chains, and the Asn-108 β → Lys mutation (rHb1.0); (2) pseudotetrameric di- α -haemoglobin with Asn-108 β → Lys and N-termini modified as already described (rHb1.1); (3) a pseudotetramer without the Asn-108 β → Lys (rHb0.1); and (4) a tetramer with neither the α -fusion nor the β -globin mutation (rHb0.0). These four forms of recombinant haemoglobin were purified from *E. coli* and analysed for functionality⁴. Table 1 summarizes these results and compares them with adult native human haemoglobin (HbA₀) purified from blood. The Val-to-Met mutation at the N termini of the α - and β -subunits (rHb0.0) results in a decrease in oxygen affinity, but introduction of the peptide linkage between the two α -chains (rHb0.1) causes an increase in oxygen affinity and a decrease in cooperativity. But introducing Asn-108 β → Lys in either the molecule with fused α -chains (rHb1.1) or non-fused α -chains (rHb1.0) results in a substantial decrease in oxygen affinity and a slight increase in the Bohr effect. Soluble haemoglobin injected into the circulatory system cannot interact with 2,3-bisphosphoglycerate (2,3-BPG) in erythrocytes and so its oxygen affinity is too high to

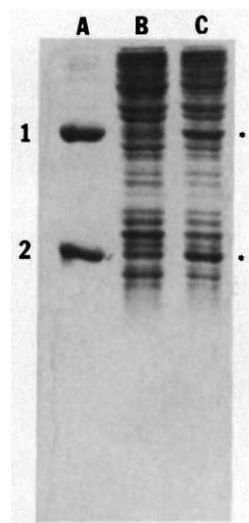


FIG. 1 Expression of rHb1.1 in *E. coli*. Lane A, 5 μ g purified rHb1.1; lane B, total cell lysate from uninduced cells bearing the plasmid pSGE1.1-E4; lane C, total cell lysate from pSGE1.1-E4-containing cells induced with 0.1 mM isopropyl β -D-thiogalactopyranoside (IPTG) for 1 h. The protein band labelled 1 corresponds to the di- α fusion protein, and 2 corresponds to the β -globin band. SDS-polyacrylamide gel electrophoresis, sample preparation and cell growth have been described⁴. Gel stained with Coomassie brilliant blue.

† To whom correspondence should be addressed.

TABLE 2 Renal function in transfused beagle dogs

Day:	Serum creatinine (μM)				Blood urea nitrogen (mg dl^{-1})				Creatine clearance ($\text{ml min}^{-1} \text{kg}^{-1}$)		NAG	
	0	2	4	8	0	2	4	8	0	2	0	2
Saline	66 $\pm$ 6	58 $\pm$ 7	62 $\pm$ 2	61 $\pm$ 5	7.1 $\pm$ 1.4	8.0 $\pm$ 2.7	12.4 $\pm$ 1.5	4.8 $\pm$ 1.1	2.5 $\pm$ 0.5	3.2 $\pm$ 0.7	39 $\pm$ 18	34 $\pm$ 24
Human serum albumin	72 $\pm$ 9	61 $\pm$ 11	61 $\pm$ 4	67 $\pm$ 7	7.3 $\pm$ 1.6	9.8 $\pm$ 0.8	11.3 $\pm$ 1.4	5.2 $\pm$ 0.8	2.3 $\pm$ 0.5	2.6 $\pm$ 1.3	42 $\pm$ 26	28 $\pm$ 17
rHb1.1	71 $\pm$ 6	59 $\pm$ 10	66 $\pm$ 4	67 $\pm$ 5	7.0 $\pm$ 1.3	7.1 $\pm$ 1.8	13.2 $\pm$ 1.8	4.9 $\pm$ 1.1	2.2 $\pm$ 0.6	3.0 $\pm$ 0.9	38 $\pm$ 18	24 $\pm$ 17
Blood (autologous)	67 $\pm$ 6	58 $\pm$ 5	61 $\pm$ 8	60 $\pm$ 11	7.0 $\pm$ 1.4	8.1 $\pm$ 2.3	11.7 $\pm$ 2.2	4.6 $\pm$ 0.9	2.3 $\pm$ 0.7	3.1 $\pm$ 0.6	44 $\pm$ 20	28 $\pm$ 13

Day-0 group values were obtained before administration of the test solution. There were 6 animals in the saline group and 8 in all other groups. Average normal values serum creatinine, 79 μM (ref. 20); blood urea nitrogen, 13 mg dl^{-1} (ref. 20); creatinine clearance, 3–4 $\text{ml min}^{-1} \text{kg}^{-1}$ (ref. 20); *N*-acetyl- β -D-glucosaminidase (NAG) activity is given in $\mu\text{mol per mmol creatinine}^{20}$. Blood (22.5 ml per kg body weight) was removed ($\sim 30\%$ of total blood volume) from four groups of anaesthetized dogs and replaced with an equal volume of normal saline with human serum albumin (50 mg ml^{-1}) and rHb1.1 (50 mg ml^{-1}), or with their own blood.

support oxygen transport. The oxygen affinity of rHb1.1 ($P_{50} = 33 \text{ mm Hg}$ at 37°C) is comparable even in the absence of 2,3-BPG to that of whole blood ($P_{50} = 28 \text{ mm Hg}$), indicating that rHb1.1 is a suitable oxygen carrier.

To determine the effect of the Asn-108 $\beta \rightarrow$ Lys mutation and the peptide linkage between two α -chains on the structure, we crystallized rHb1.1 in the deoxy form and determined its structure. Figure 2 shows a difference Fourier map of rHb1.1 minus deoxy HbA in the C-terminal region of one α -subunit. The peptide linkage breaks the molecular dyad symmetry of the haemoglobin tetramer, but this asymmetry is not sufficient to orient tetramers uniquely in the crystal. Therefore one configuration with the free N and C termini of the α 1-Gly- α 2 polypeptide

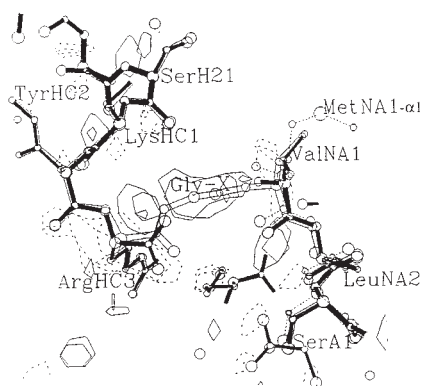


FIG. 2 Symmetry-averaged difference map of rHb1.1 minus HbA superimposed on the atomic model. The map is contoured at plus (solid contours) and minus (broken contours) about twice the root-mean-square value of the electron density before symmetry-averaging about the molecular dyad. Filled bonds represent the refined coordinates of deoxy HbA (ref. 11). Open bonds and single dotted bonds represent approximate coordinates of rHb1.1 in the region of one α 1-Gly- α 2 linkage: open bonds show the orientation with the link, and single bonds the dyad-related orientation with free N and C termini.

METHODS. rHb1.1 was crystallized from high salt in space group $P2_1$ (ref. 22). X-ray data to 2.5 Å resolution were collected from a single crystal on a FAST/MADNES area detector with a rotating anode source. Absorption (calculated from variation of the direct beam with crystal orientation), Lorentz and polarization corrections, and batch scale and temperature factors (batches of 5° rotation in angle ϕ), were applied to the raw X-ray data, which were then merged to yield 18,231 unique reflections, or 96% of the total between 10 and 2.5 Å resolution. Anomalous dispersion measurements were made for 76% of unique reflections; the *R*-factor on intensity between Friedel pairs in the reduced data set was 6.6% on intensity. Reduced data were converted to amplitudes and scaled to native deoxy HbA data with an *R*-factor of 16%. The difference map was calculated with phases obtained from the atomic model of deoxy HbA (ref. 11) and averaged about the molecular dyad axis. The similarity of the unaveraged difference map in the two N- and C-terminal regions indicates that the two alternative configurations are present at nearly equal occupancies.

is superimposed on the dyad-related configuration with Gly-linked α 1 C terminus and α 2 N terminus. Positive densities bridging the N and C termini indicate that the glycine residue links the C terminus of an α 1 chain with the N terminus of the dyad-related α 2 chain. The negative and positive densities around a few residues near both N and C termini indicate that the peptide linkage between the two α -subunits slightly constrains the conformation of both termini. The Asn-108 $\beta \rightarrow$ Lys mutation causes substantial structural changes, including a shift of the B and G helices indicated by paired positive and negative densities along these two helices (not shown), but the structural basis of the reduced affinity is not obvious.

To test whether the fusion of α -subunits reduces renal filtration and extends the intravascular half-life of haemoglobin, we prepared radioactively labelled rHb1.1 and rHb1.0 for investigation in rats. On the basis of experiments with chemically cross-linked haemoglobin^{17–19}, we expected to see a 2–4-fold increase in the intravascular half-life of rHb1.1 compared with rHb1.0 at the dose given. As shown in Fig. 3, the half-life of rHb1.1

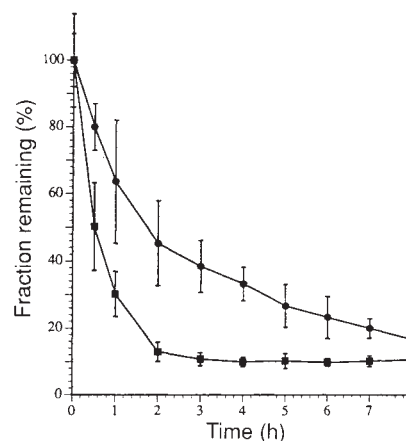


FIG. 3 *In vivo* half-life of rHb1.1 (circles) and rHb1.0 (squares). The fraction of ^3H -rHb1.0 or rHb1.1 remaining in the intravascular compartment is shown on the y axis. The value for 100% was determined from samples taken 5 min after administration of the haemoglobin samples.

METHODS. Radioactive haemoglobin was prepared by adding 5 mCi ^3H -leucine (Amersham; 1,100–1,500 Ci mmol $^{-1}$) to 2-l fermenters of *E. coli* JM109[pSGE1.1-E4] or JM109[pSGE1.0-E4] 15 min before induction with IPTG. Haemoglobin was purified essentially as described⁴ and combined with unlabelled rHb1.1 or rHb1.0 to adjust the specific activity to 8.3×10^3 d.p.m. mg^{-1} (rHb1.1) and 9.6×10^3 d.p.m. mg^{-1} (rHb1.0). Eight Sprague-Dawley rats (200–300 g) in 2 groups of 4 were given either 1 ml ^3H -rHb1.1 (145 mg ml^{-1}) or 0.9 ml ^3H -rHb1.0 (156 mg ml^{-1}) through a venous catheter. Blood samples (0.2 ml) were removed at 5 min, 30 min, and at 1, 2, 3, 4, 5, 6, 7 and 8 h from an arterial catheter. Animals were anaesthetized with a mixture of ketamine hydrochloride and xylazine by intramuscular injection before insertion of catheters and allowed to recover for three days before administration of haemoglobin solutions.

containing fused α -subunits is increased 3–4-fold compared with its non-fused homologue. We also studied the effects of rHb1.1 on renal function in dogs, removing 30% of the blood volume and replacing it with an equivalent volume of saline, 50 mg ml⁻¹ human serum albumin, 50 mg ml⁻¹ rHb1.1, or with the animal's own blood. Serum creatinine, blood urea nitrogen, creatinine clearance rate and excretion of *N*-acetyl- β -D-glucosaminidase were measured as indicators of glomerular function: serum creatinine and urea nitrogen increase and creatinine clearance is slower if glomerular function is compromised. Increased *N*-acetyl- β -D-glucosaminidase in the urine is a sensitive measure of papillary damage in the kidney of

dogs²⁰. The results in Table 2 show that kidney function was normal in all groups. Kidney histology was also normal in all groups at 7 or 14 days after administration of the test solution (data not shown), and no haemoglobin was detected in the urine of the animals (limit of detection is 1.5 μ g ml⁻¹; data not shown). These results are in contrast to the dramatic nephrotoxicity of unmodified haemoglobin^{8,9,21}.

Unlike haemoglobins that have been chemically crosslinked, genetically fused di- α -haemoglobin can be produced in large amounts by simple microbial fermentation and purified without further modification. Our engineered haemoglobin is therefore a strong candidate for a safe blood substitute. □

Received 12 September 1991; accepted 6 January 1992.

1. Pool, R. *Science* **250**, 1654–1656 (1990).
2. Moss, G. S., Gould, S. A., Sehgal, H. L., Sehgal, L. R. & Rosen, A. L. *Surgery* **95**, 249–255 (1984).
3. Mulder, A. G., Amberson, W. R., Steggerda, F. R. & Flexner, J. *J. cell. comp. Physiol.* **5**, 383–397 (1934).
4. Hoffman, S. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 8521–8525 (1990).
5. Wagenbach, M. *et al. Biotechnology* **9**, 57–61 (1991).
6. Benesch, R. & Benesch, R. E. *Biochem. biophys. Res. Commun.* **26**, 162–167 (1967).
7. Ackers, G. K. & Halvorson, H. R. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4312–4316 (1974).
8. Bunn, H. F., Esham, W. & Bull, R. J. *J. exp. Med.* **129**, 909–923 (1969).
9. Bunn, H. F. & Jandl, J. *J. exp. Med.* **129**, 925–934 (1969).
10. Lee, R., Atsumi, N., Jacobs, E., Austen, W. & Vlahakes, G. *J. Surg. Res.* **47**, 407–411 (1989).
11. Fermi, G., Perutz, M., Shaanan, B. & Fourme, R. *J. molec. Biol.* **175**, 159–174 (1984).
12. Shaanan, B. *J. molec. Biol.* **171**, 31–59 (1983).
13. Nagai, K., Enoki, Y., Tomita, S. & Teshima, T. *J. biol. Chem.* **257**, 1622–1625 (1982).
14. Better, J., Chang, P., Robinson, R. & Horwitz, R. *Science* **240**, 1041–1043 (1988).
15. Skerra, A. & Pluckthorn, A. *Science* **240**, 1038–1041 (1988).
16. Moo-Penn, W. *et al. FEBS Lett.* **92**, 53–57 (1978).
17. Bucci, E., Razynska, A., Urbaitis, B. & Fronticelli, C. *J. biol. Chem.* **264**, 6191–6195 (1989).
18. Snyder, S. R., Welty, E. V., Walder, R. Y., Williams, L. A. & Walder, J. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7280–7284 (1987).
19. Manning, L. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 3329–3333 (1991).
20. Loeb, W. F. & Quimby, F. W. (eds) *The Clinical Chemistry of Laboratory animals* 460–495 (Pergamon, New York, 1989).
21. Biro, G. P., Ning, J., Peterson, L. M. N. & Anderson, P. J. *Biomater. Artificial Cells Immob. Technol.* **19**, 357 (1991).
22. Perutz, M. F. *J. Crystal Growth* **2**, 54–56 (1968).

ACKNOWLEDGEMENTS. We thank M. Perutz and G. Dodson for critical comments on this manuscript, C. Scoggins for discussion and K. Morris for help with animal studies.

Role of ATP and disulphide bonds during protein folding in the endoplasmic reticulum

Ineke Braakman, Jonne Helenius & Ari Helenius

Department of Cell Biology, Yale University School of Medicine,
333 Cedar Street, New Haven, Connecticut 06510-8002, USA

BEING topologically equivalent to the extracellular space, the lumen of the endoplasmic reticulum (ER) provides a unique folding environment for newly synthesized proteins. Unlike other compartments in the cell where folding occurs, the ER is oxidizing and therefore can promote the formation of disulphide bonds¹. The reducing agent dithiothreitol, when added to living cells, inhibits disulphide formation with profound effects on folding². Taking advantage of this effect, we demonstrate here that folding of influenza haemagglutinin is energy dependent. Metabolic energy is required to support the correct folding and disulphide bond formation in this well characterized viral glycoprotein, to rescue misfolded proteins from disulphide-linked aggregates, and to maintain the oxidized protein in its folded and oligomerization-competent state.

To monitor the folding process, we used a previously developed pulse-chase assay in influenza-infected cells³. At the end of each chase time, the cells were treated with a membrane permeant alkylating agent, *N*-ethyl maleimide (NEM), to trap folding intermediates⁴. The formation of disulphide bonds was followed by the changing mobility of haemagglutinin (HA0) bands in nonreducing SDS-PAGE. Conformational comparisons were made by immunoprecipitation with a panel of monoclonal and polyclonal antibodies.

The post-translational folding and maturation of HA0 in unperturbed CHO15B cells is shown in Fig. 1. After a 2-min pulse, two major folding intermediates, IT1 and IT2, were visible as well as the fully oxidized, untrimmed HA0 (called NT)³. During the chase, IT1 and IT2 were converted to NT, which then trimerized (not shown) and moved to the Golgi complex where its carbohydrates were trimmed, resulting in a faster moving band marked G (for Golgi) (Fig. 1, lanes 4–6). When

reduced, IT1, IT2, and NT migrated as a single band (R) with a mobility slightly lower than IT1. The trimmed G (GR) was resolved again as a faster moving band (lane 7). When dithiothreitol (DTT) was added to the chase medium, the radioactively labelled IT1, IT2 and NT were reduced. The reduced form R appeared whether the cells were exposed to ATP depletion or

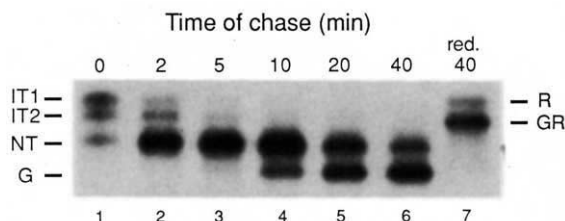


FIG. 1 Post-translational folding of HA0 in CHO15B cells. Two folding intermediates with an immature set of disulphide bonds (IT1 and IT2) gradually folded into the native form (NT), which has all 6 disulphide bridges. Transport out of the ER is apparent from the increase in mobility (G), caused by trimming of carbohydrates from Man₅GlcNAc₂Asn to Man₅GlcNAc₂Asn. When reduced, IT1, IT2 and NT ran as one band (R) (lane 7). GR denotes the reduced Golgi-form of HA0. The CHO15B cells were used in our experiments because they are defective in *N*-acetylglucosaminyltransferase^{29,30}, and the trimmed Man₅GlcNAc₂Asn form of the *N*-linked carbohydrates cannot be processed further. This allows easy distinction between pre-Golgi (NT) and later forms (G) of HA0 by SDS-PAGE.

METHODS. Folding of HA0 was monitored as described before³. Briefly, 6 h after infection with X31 influenza virus (derived from A/Aichi/1968/H3N2), CHO15B cells were pulse-labelled for 1 min with 125 μ Ci ml⁻¹ each of [³⁵S]-cysteine and Tran[³⁵S]-label. A chase with 5 mM methionine and 5 mM cysteine followed, with 0.5 mM cycloheximide in the medium to block elongation of labelled nascent chains. Ice-cold PBS with 20 mM *N*-ethylmaleimide stopped the chase and prevented rearrangement of disulphide bonds. This concentration of *N*-ethylmaleimide was at least 10-fold the concentration needed to consistently preserve IT1 and IT2. The cells were lysed in ice-cold 0.5% Triton X-100 in MNT (20 mM MES, 100 mM NaCl, 30 mM Tris-HCl, pH 7.5) containing 20 mM *N*-ethylmaleimide, 1 mM EDTA, 1 mM PMSF and 10 μ g ml⁻¹ each of chymostatin, leupeptin, antipain and pepstatin. Lysates were immunoprecipitated with a polyclonal rabbit antiserum against X31 virus (P), which recognizes all forms of the haemagglutinin^{3,7,8}. Reducing and nonreducing SDS-PAGE (7.5%), followed by fluorography, was used to analyse the folding process.

not (Fig. 2, lanes 2 and 5). The ATP-depleting medium used was glucose-free and contained sodium azide and 2-deoxy-D-glucose^{5,6}. When the DTT was washed out, rapid reoxidation occurred. In control cells, IT1 and IT2 reappeared, NT was formed (lanes 3, 4), and the protein was transported to the Golgi complex.

In energy-depleted cells, IT1, IT2 and NT were only seen transiently (lanes 6, 7) before HA0 moved into aggregates on top of the gel. The aggregates increased in size with time, resulting in material that would not enter the stacking gel. If the samples were reduced before electrophoresis, the labelled HA0 in the aggregates was quantitatively recovered as a reduced, nontrimmed monomer (R), indicating that the HA0 in the energy-depleted cells was not being degraded, that it was confined to the ER, and that it was present in disulphide cross-linked complexes. When DTT was already present during translation, HA0 was synthesized in a completely reduced form (R; Fig. 2, lane 10). After DTT wash-out under energy-depleting conditions (Fig. 2, lanes 12–15), it also aggregated, but without the transient formation of IT1, IT2 or NT, indicating that the fully reduced HA0 could not fold properly in energy-depleted cells.

To determine whether folding during normal HA0 biosynthesis also required metabolic energy, we pulse-labelled cells and then depleted them of ATP during the chase. DTT was not added at any stage of the experiment. Again, disulphide-linked

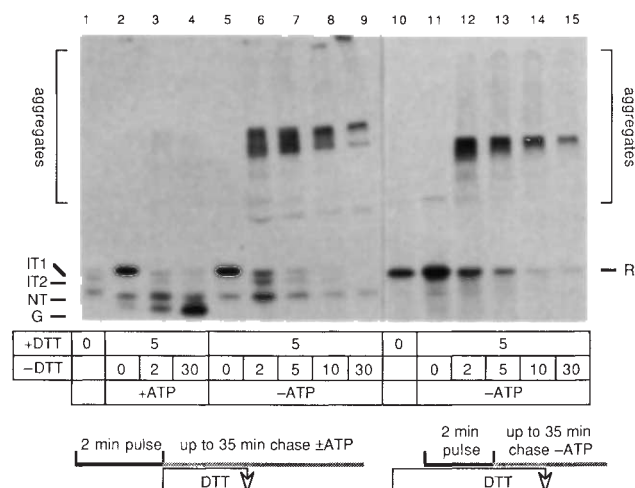


FIG. 2 ATP dependence of disulphide-bond formation. CHO15B cells were pulse-labelled for 2 min. DTT (5 mM) was then added in the chase. Post-translational reduction of HA0 was very fast *in vivo* (lane 2). When DTT was washed out in the presence of ATP, only a very small part of the protein was trapped into an aggregate by aberrant interchain disulphide bond formation (lane 3). These aggregates were dimers of HA0, trimers, and larger clusters, and they did not change over time (lane 4). Reoxidation in the absence of ATP, however, landed the major part of HA0 in disulphide-linked aggregates (lanes 6–9), which gradually increased in size. When reduced before SDS-PAGE, HA0 ran as one band (not shown). None of the HA0 was allowed to exit the ER. Lanes 10, 11 show HA0 synthesized in the presence of 5 mM DTT. When DTT was now washed out in the absence of ATP (lanes 12–15), the same aggregates appeared as in lanes 6–9, also increasing in size. In contrast to this, however, no oxidized forms of HA0 appeared at any time.

METHODS. Procedures were essentially as described in Fig. 1. Lanes 1–9, 2-min pulse followed by a chase with 5 mM DTT to reduce HA0 *in vivo*. The reducing agent was then removed by two changes of chase medium without DTT followed by an incubation of up to 30 min. In lanes 10–15, the DTT was also added 5 min before and during the pulse. Care was taken that all cells received the same concentration of DTT for the same time before the pulse started. In lanes 5–15, the cells were depleted of ATP during the chases with and without DTT. The ATP-depleting system consisted of glucose-free DMEM with 20 mM 2-deoxy-D-glucose and 10 mM sodium azide.

TABLE 1 Conformation of various forms of HA0 determined by immunoprecipitation

	F1	A1	anti-BiP
HA0 intermediates in the unperturbed cell			
IT1	+++	–	–
IT2	–	–	–
NT	–	–	–
G	–	–	–
HA0 after translation in the presence of DTT			
R (no chase)	+++	–	–
R (5' chase + DTT – ATP)	+++	+++	+
HA0 aggregates after wash-out of DTT in the absence of ATP			
after translation + DTT	+++	+++	+
after post-translational reduction	+++	+++	+

Detergent lysates containing the different forms of radioactively labelled HA0 were used for immunoprecipitation with the conformation-specific antibodies F1 and A1 against HA0 and with the monoclonal anti-BiP antibody. Immunoprecipitations were done on detergent lysates (Fig. 1) as described³. The antibodies used were from the following sources: F1 is a mouse monoclonal IgG made against nondenatured HA2 (ref. 3), and A1 is a mouse monoclonal antibody made against acid-treated HA0, which recognizes misfolded forms of HA0^{7,8}. The monoclonal antibody to BiP precipitates native BiP¹⁰ as well as complexes between BiP and misfolded HA0^{7,10–13,28}. The fraction of HA0 recognized by the antibodies was roughly estimated and compared with the amount immunoprecipitated by the polyclonal antiserum P (=100%): +, <40%; +++, >80%. IT1, IT2, NT and G are defined as in Fig. 1, and R stands for any reduced form of HA0.

aggregates were formed (Fig. 3a, lanes 2–5). IT1 and IT2, as well as NT formed during the pulse, disappeared and moved into the covalently linked aggregates, indicating that already folded proteins were modified and trapped into aggregates when the energy level dropped. We also tested what would happen if cells were allowed to regenerate ATP after the formation of aggregates. The results in Fig. 3a, b (lanes 6, 7 and 5, 6, respectively) showed that a large fraction of the labelled HA0 returned to a monomeric form and proceeded to fold normally. The misfolded molecules in the disulphide crosslinked aggregates were thus rescued when metabolic energy was made available again.

The aggregates observed in the various energy-depleted cells represented misfolded HA0. In addition to having aberrant interchain disulphide bonds, they were precipitated by monoclonal antibody A1 which is specific for misfolded and acid-treated HA0^{7,8}, and by antibody F1 which sees the first folding intermediate IT1 but not IT2, normally folded monomers (NT) or mature HA0 trimers (Table 1). The major top domain epitopes A, B and E, present in all normal forms of HA0, were only marginally expressed (not shown). Moreover, the aggregates coprecipitated with BiP (or grp 78)^{9,10}, a property of a variety of misfolded forms of HA0⁷ and other proteins^{11–13}. Although the anti-BiP antibody precipitates folding intermediates of numerous other proteins, it did not precipitate any of the normal maturation intermediates of X31 HA0 (Table 1). The antibody precipitation experiments indicated that misfolding also occurred when ATP was depleted in the presence of DTT (Table 1, line 6), that is under conditions where no disulphides could form. The misfolded conformation after depletion of metabolic energy was, therefore, not directly dependent on either the presence of aberrant or the absence of native disulphide bonds.

Our results revealed that metabolic energy was required to secure the correct folding of newly synthesized HA0, to rescue misfolded HA0 molecules from disulphide crosslinked aggregates, and to maintain already folded monomeric HA0 in a folded state. They suggested that HA0 in the ER is subject to

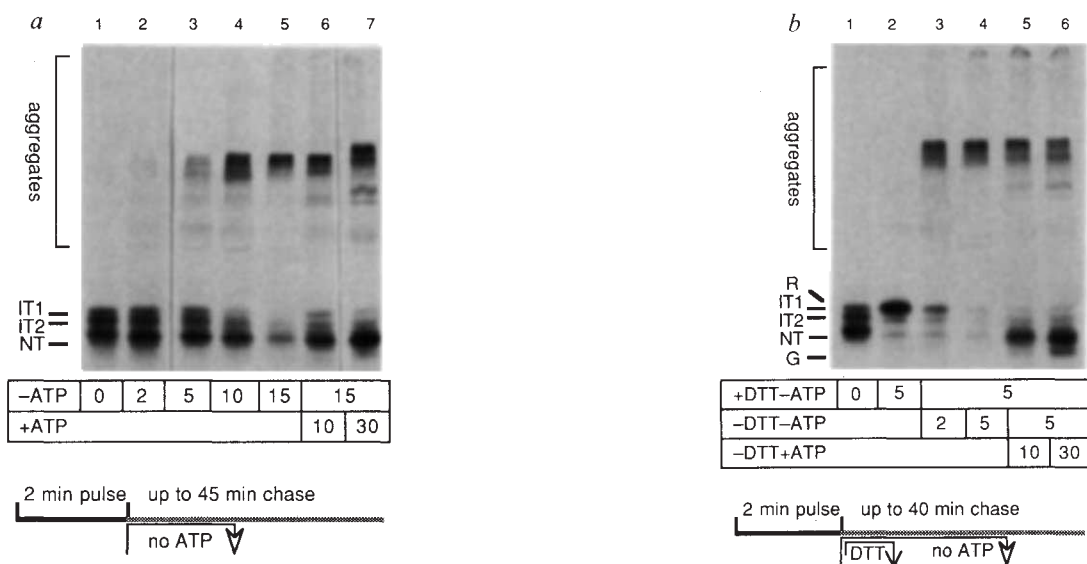


FIG. 3 ATP dependence of HAO folding and refolding. *a*, When ATP was depleted during a chase immediately after pulse-labelling, HAO formed disulphide-linked aggregates, which increased in size during the chase. The slow onset of this aggregate formation was most probably caused by the delay in the ATP depletion. The aggregates were similar to the ones observed after previous reduction of HAO (Fig. 2, lanes 6–9). When cells were allowed to recover from the ATP depletion, a large fraction of the HAO returned from the aggregates into folding intermediates, folded into NT (lanes 6–7), trimerized, and was transported to the Golgi. *b*, After ATP depletion during post-translational reduction and reoxidation, cells were allowed to recover

(lanes 5, 6). Aggregates decreased in size, and HAO reappeared on the gel as IT1, IT2 and NT. After a longer recovery time, the correctly folded NT was trimerized and transported, as judged by the presence of G (lane 6). **METHODS.** *a*, Conditions were as in Fig. 1, with one exception: the chase medium consisted of glucose-free DMEM with 20 mM 2-deoxy-D-glucose and 10 mM sodium azide to deplete ATP in the cells. After 15 min, this medium was replaced with regular medium to allow recovery. *b*, Conditions were as in Fig. 2, lanes 5–9. After 10 min total chase, the ATP-depleting medium was replaced by regular chase medium by two washes, to allow the cells to recover from the ATP depletion.

continuous folding, unfolding and refolding with energy required for folding and refolding. Because HAO monomers (NT) depended on energy to remain oxidized and correctly folded, they apparently did not reach a stable minimum-energy-state. Such a state is probably attained only when the HAO trimerizes and reaches the Golgi (U. Tatu, I.B., J.H. and A.H., unpublished result). We conclude that the ER lumen in the energized cell is a highly dynamic folding environment.

Protein folding in mitochondria and in the cytosol of bacteria is energy-dependent^{14–17}. ATP is hydrolysed by chaperones that associate with the folding polypeptides and prevent nonproductive intra- and intermolecular interactions. For the ER, the situation has been less clear. A misfolded, temperature-reversible folding mutant, the ts045 G protein of vesicular stomatitis virus, required ATP for its refolding after a shift from nonpermissive to permissive temperature¹⁸, but we could not demonstrate a similar requirement for the wild-type protein. MHC class I antigen assembly *in vitro* is energy-dependent¹⁹. Evidence that ATP is present in the ER lumen came with the discovery of an ATP translocator in the ER membrane²⁰.

The most likely molecular explanation for the energy requirement is the involvement of BiP²¹ in the post-translational folding process. BiP is an hsp70 analogue²² generally assumed to be a folding factor in the ER^{23–25}. It binds and hydrolyses ATP as it associates and dissociates from peptides and proteins^{22,26}. The involvement of BiP is consistent with the observation that CCCP (carbonyl cyanide *m*-chlorophenyl hydrazone) prevents secretion of factor VIII and induces its accumulation in a complex with BiP²⁷. We have been unable to detect BiP binding to normal folding intermediates of HAO (Table 1), but this could simply mean that the associations are rapidly reversed or too ATP-sensitive to withstand the immunoprecipitation conditions. Another explanation for the observed energy-dependence of folding could be that energy is required to maintain oxidizing conditions inside the ER. We believe this not to be the case because ATP depletion did not

lead to inhibition of disulphide bond formation, but rather to the formation of incorrect disulphide bonds. Apparently the oxidizing conditions in the ER were maintained (and re-established after DTT wash-out) in energy-depleted cells. We also observed HAO misfolding during ATP-depletion even in the presence of DTT, that is, independent of disulphide bonds.

Received 23 September; accepted 12 December 1991.

1. Freedman, R. B. *Cell* **57**, 1069–1072 (1989).
2. Braakman, I., Helenius, J. & Helenius, A. *EMBO J.* (in the press).
3. Braakman, I., Hoover-Litty, H., Wagner, K. R. & Helenius, A. *J. Cell Biol.* **114**, 401–411 (1990).
4. Creighton, T. E. *Prog. Biophys. Molec. Biol.* **33**, 231–297 (1978).
5. Clarke, B. L. & Weigel, P. H. *J. Biol. Chem.* **260**, 128–133 (1985).
6. Podbilewicz, B. & Mellman, I. *EMBO J.* **9**, 3477–3487 (1990).
7. Hurlley, S. M., Bole, D. G., Hoover-Litty, H., Helenius, A. & Copeland, C. S. *J. Cell Biol.* **108**, 2117–2126 (1989).
8. Copeland, C. S., Doms, R. W., Bolzau, E. M., Webster, R. G. & Helenius, A. *J. Cell Biol.* **103**, 1179–1191 (1986).
9. Haas, I. G. & Wabl, M. *Nature* **306**, 387–389 (1983).
10. Bole, D. G., Hendershot, L. M. & Kearney, J. F. *J. Cell Biol.* **102**, 1558–1566 (1986).
11. Dörner, A. J., Bole, D. G. & Kaufman, R. J. *J. Cell Biol.* **105**, 2665–2674 (1987).
12. Kassenbrock, C. K., Garcia, P. D., Walter, P. & Kelly, R. B. *Nature* **333**, 90–93 (1988).
13. Kozutsumi, Y., Segal, M., Normington, K., Gething, M.-J. & Sambrook, J. *Nature* **332**, 462–464 (1988).
14. Bochkareva, E. S., Lissin, N. M. & Girshovich, A. S. *Nature* **336**, 254–257 (1988).
15. Laminet, A. A., Ziegelhoffer, T., Georgopoulos, C. & Plückthun, A. *EMBO J.* **9**, 2315–2319 (1990).
16. Martin, J. et al. *Nature* **352**, 36–42 (1991).
17. Ellis, R. J. & Van der Vies, S. M. A. *Rev. Biochem.* **60**, 321–347 (1991).
18. Doms, R. W., Keller, D. S., Helenius, A. & Balch, W. E. *J. Cell Biol.* **105**, 1957–1969 (1987).
19. Lévy, F., Gabathuler, R., Larsson, R. & Kvist, S. *Cell* **67**, 265–274 (1991).
20. Clairmont, C., Demaio, A. & Hirschberg, C. *J. Cell Biol.* **115**, 255a (1991).
21. Kassenbrock, C. K. & Kelly, R. B. *EMBO J.* **8**, 1461–1467 (1989).
22. Munro, S. & Pelham, H. R. B. *Cell* **46**, 291–300 (1986).
23. Gething, M.-J. & Sambrook, J. *Semin. Cell Biol.* **1**, 65–72 (1990).
24. Rothman, J. E. *Cell* **59**, 591–601 (1989).
25. Pelham, H. R. B. *Cell* **46**, 959–961 (1986).
26. Flynn, G. C., Chappel, T. G. & Rothman, J. E. *Science* **245**, 385–390 (1989).
27. Dörner, A. J., Wasley, L. C. & Kaufman, R. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7429–7432 (1990).
28. Gething, M.-J., McCammon, K. & Sambrook, J. *Cell* **46**, 939–950 (1986).
29. Gottlieb, C., Baenziger, J. & Kornfeld, S. *J. Biol. Chem.* **250**, 3303–3309 (1975).
30. Balch, W. E., Elliott, M. M. & Keller, D. S. *J. Cell Biol.* **261**, 14681–14689 (1986).

ACKNOWLEDGEMENTS. We thank D. Bole for the anti-BiP antibody, H. Tan for photography, and the Helenius-Mellman laboratory for advice and criticism. This work was supported by grants from the NIH to A.H. I.B. was supported by a post-doctoral fellowship from the Human Frontier Science Program Organization.

A protein-folding reaction under kinetic control

David Baker, Julie L. Sohl & David A. Agard

Howard Hughes Medical Institute, The Department of Biochemistry and the Biophysics Graduate Group, University of California at San Francisco, San Francisco, California 94143-0448, USA

SYNTHESIS OF α -lytic protease as a precursor containing a 166 amino-acid pro region¹ transiently required for the correct folding of the protease domain²⁻⁴. By omitting the pro region in an *in vitro* refolding reaction we trapped an inactive, but folding competent state (I) having an expanded radius yet native-like secondary structure. The I state is stable for weeks at physiological pH in the absence of denaturant, but rapidly folds to the active, native state on addition of the pro region as a separate polypeptide chain. The mechanism of action of the pro region is distinct from that of the chaperonins^{5,6}; rather than reducing the rate of off-pathway reactions, the pro region accelerates the rate-limiting step on the folding pathway by more than 10^7 . Because both the I and native states are stable under identical conditions with no detectable interconversion, the folding of α -lytic protease must be under kinetic and not thermodynamic control.

Previously, we found that α -lytic protease denatured in 6 M guanidine under nonreducing conditions regained enzymatic activity when the guanidine was removed by dialysis in the presence but not in the absence of the pro region⁴. That the pro region strongly inhibits the native enzyme (inhibition constant, $K_i \sim 10^{-10}$ M (ref. 4)) suggested that it functioned in a late step in the folding pathway. Thus, we reasoned that it might be possible to trap a folding intermediate by removing denaturant in the absence of the pro region. Removal of the guanidine by dialysis resulted in considerable sample aggregation and recovery of less than 4% of the starting activity after addition of the pro region. By contrast, little aggregation occurred when the guanidine concentration was rapidly lowered by dilution. The addition of stoichiometric amounts of a glutathione transferase-pro region fusion protein (GEX-PRO, see ref. 4) to the diluted material resulted in recovery of 74% of the input activity

TABLE 1 Dependence of folding competence on ionic strength and temperature

	Activity (% $t=0$)	Absorbance at 340 nm
(a) Tris concentration (mM)		
5	75	0.008
50	40	0.020
100	26	0.053
200	15	0.094
500	3	0.142
(b) Temperature		
4	85	<0.001
24	3.5	0.21
37	0.25	0.23

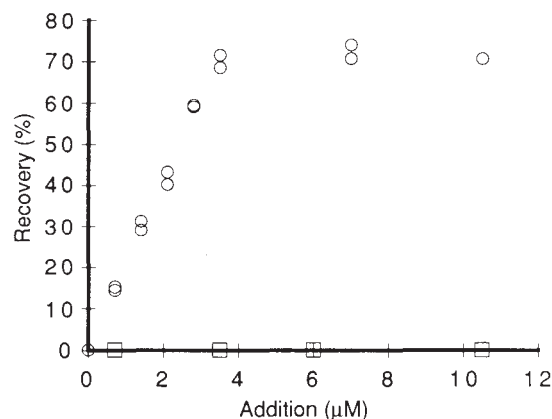
To investigate the stability of the I state, denatured α -lytic protease was diluted into buffers of varying ionic strength, incubated for 16 h, and then supplemented with GEX-PRO and assayed for refolding. Recovery of activity decreased with increasing ionic strength and increasing temperature. This decrease correlated with an increase in turbidity (A_{340}) suggesting that aggregation accounted for the loss in refolding competence in both cases. The decrease in aggregation at low ionic strength is likely to result from the increased electrostatic repulsion of the highly charged monomers ($pI = 10.2$) due to the loss of counter-ion shielding. The increased stability at low temperature may result from the reduction in the strength of hydrophobic interactions that are likely to mediate aggregation. **Methods.** a, Denatured α -lytic protease in 6 M guanidine HCl was diluted (final concentration of protease 2.5 μ M) into the indicated concentration of Tris-Cl plus 100 μ M AcProBoro Val and incubated for 16 h at 24 °C. Samples were then supplemented with a twofold molar excess of GEX-PRO and assayed for recovery of enzymatic activity as described in Fig. 1 legend. The absorbance of the solution at 340 nm was also monitored to assess the degree of aggregation of the protease. b, Denatured α -lytic protease was diluted into 5 mM KPO₄ pH 8.0, 100 μ M AcProBoroVal (final concentration of protease 6 μ M) and incubated for 16 h at the indicated temperature. The diluted protease solutions were then treated as in a.

whereas GEX alone had no effect (Fig. 1). This inactive, but folding competent, state of the protease (I) is either directly on the folding pathway or in rapid equilibrium with a conformation on the folding pathway.

A large energy barrier must block the conversion of the I state to the native state in the absence of the pro region. To estimate

FIG. 1 Specificity and stoichiometry of pro region-dependent conversion of the I state to the native state. The inactive but folding competent state I was prepared by rapid dilution as described below and incubated with either pro region supplied as a glutathione transferase-pro region fusion (GEX-PRO⁴, $\circ$) or glutathione transferase alone (GEX, $\square$) as a control. Samples were treated with trypsin to relieve pro-region inhibition and assayed for recovery of protease activity. Trypsin rapidly destroys the pro region and unfolded α -lytic protease species, while leaving the folded enzyme intact⁴. The concentration of the I state determined from the absorbance at 280 nm was 4.4 μ M. The maximum amount of native protease recovered was 3.25 μ M corresponding to a 74% recovery. The refolding incompetent material may correspond to proteolytic fragments not removed by dialysis or to small aggregates. The observed requirement for stoichiometric amounts of GEX-PRO is probably due to strong product inhibition; the pro region binds the folded enzyme with extremely high affinity ($K_i \sim 10^{-10}$ M (ref. 4)).

METHODS. α -Lytic protease was purified essentially as described in ref. 9, except that the cation exchange chromatography step was done on S-Sepharose at pH 9.6 with a 50–200 mM salt gradient for elution. Enzyme was dissolved at 200 mg ml⁻¹ in 100 mM Tris-Cl pH 8.0 and then denatured by addition of 3 volume equivalents of 8 M guanidine-Cl, 0.67 M glycine pH 3.0. This procedure minimized autolysis at high protease concentrations. The protease was incubated in the 6 M guanidine solution for at least 4 h at 24 °C to ensure complete denaturation. The denatured protein was extensively dialysed (using a 6–8 K cutoff membrane) against 6 M guanidine to remove any fragments from residual autolysis during denaturation. Dilution was carried out by placing a small (5–10 μ l) portion of denatured protease solution at the bottom of a test tube and rapidly dispersing the



drop by addition of a 200–500-fold excess of buffer lacking guanidine. The dilution buffer contained 100 μ M AcProBoroVal, a low-affinity ($K_i = 3$ μ M (ref. 13)) α -lytic protease inhibitor, to prevent possible autolysis by residual folded protease. Aliquots of the diluted protease solution were supplemented with GEX or GEX-PRO proteins as indicated and incubated for at least 15 min at 4 °C. To relieve inhibition by the pro region, samples were then treated with trypsin (5 μ g) for 10 min at 24 °C. α -Lytic protease activity was then assayed as described in ref. 2. Trypsin does not interfere with the α -lytic protease assay⁴.

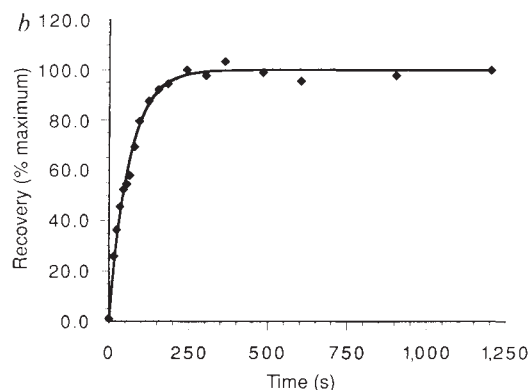
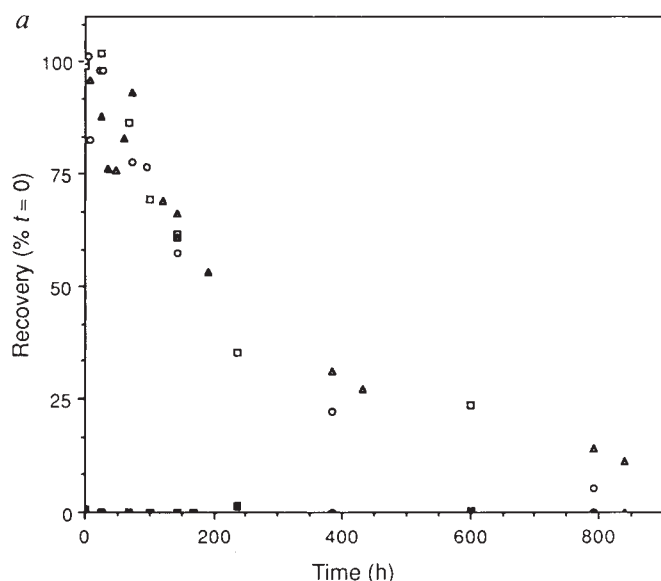


FIG. 2 Kinetics of folding in the absence and presence of the pro region. *a*, To estimate the height of the barrier blocking the folding of the I state to the active enzyme, the I state was prepared as described in Fig. 1, adjusted to a final concentration of 0.3 μM ($\square$), 1 μM ($\circ$) or 5 μM ($\triangle$) and incubated at 4 $^{\circ}\text{C}$ for the indicated time. Samples were then supplemented with a twofold molar excess of GEX (filled symbols) or GEX-PRO (open symbols) and processed as described in Fig. 1. The slow loss of refolding competence

with time may result from aggregation, proteolysis, disulphide exchange or other chemical modification. The detection limit in the protease assay is roughly 0.3 nM, corresponding to a recovery of 0.3% of the starting activity in the most concentrated sample (5 μM). *b*, The kinetics of folding in the presence of the pro region were measured by mixing the I state (0.4 μM) with GEX-PRO (0.04 μM) at 4 $^{\circ}\text{C}$ at $t=0$ and adding trypsin (final concentration 0.7 mg ml^{-1}) at the indicated times to destroy the pro region and any unfolded α -lytic protease. After 10 min at 24 $^{\circ}\text{C}$, α -lytic protease activity was assayed. Extrapolation of the data to zero time suggests that the lag time for proteolysis by trypsin was <3 s. A plot of a simple exponential fit to the kinetic data is shown (solid line, $k=0.016 \text{ s}^{-1}$). The half time of folding was independent of the intermediate concentration over a 20-fold range (0.4–8 μM) indicating that the apparent dissociation constant of the I state for the pro region is below 0.4 μM .

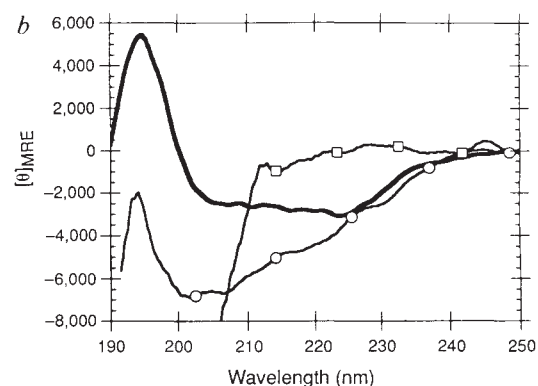
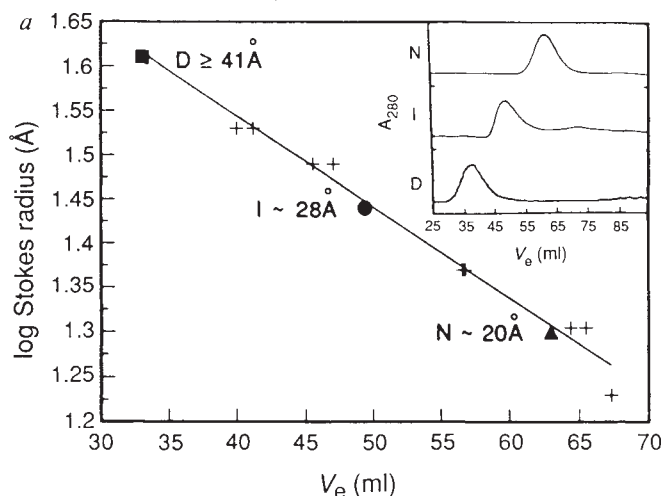


FIG. 3 Structural characterization of the I state. *a*, Gel filtration traces (shown in insert) for the I state, native (N) and denatured (D) forms of α -lytic protease. The hydrodynamic radius of the I state is intermediate between the native and denatured species. The crosses (main figure) indicate positions of molecular mass calibration standards. *b*, Peptide-region CD spectra of N, I and D. The secondary structure content of the I state was estimated by spectral deconvolution:

	% α helix	% β sheet	% β turn	% unordered
Native (bold)	4	75	2	19
I-state (circles)	0	70	3	28
Denatured (squares)	0	6	41	53

Although CD provides a poor measure of β -sheet content, the estimates obtained by spectral deconvolution of the peptide region agree well with values calculated from the X-ray crystal structure¹⁴ using the Kundrot and Richards algorithm¹⁵: 74% β sheet and 2.5% α helix.

METHODS. *a*, A G-75 Sephadex column (3–80K fractionation range) at 4 $^{\circ}\text{C}$ was used for all sizing data. This resin was selected because it showed minimal interaction with native protease under the low salt conditions required for stability of the intermediate. The degree of interaction of the expanded intermediate with the gel matrix is unknown. Therefore, the apparent molecular mass for the I state must be considered a minimum

estimate. The running buffer was 10 mM Tris, 40 mM NaCl pH=8.0 for both N and I. Denatured dialysed protease was prepared as described in Fig. 1 and diluted to 30 mM (I) or 1 M guanidine (D). The 1 M guanidine sample (D) was run in 10 mM Tris, 1 M guanidine-HCl pH=8.0. Higher concentrations of guanidine were not used owing to low flow rates at 4 $^{\circ}\text{C}$, however CD indicates that the I state is denatured in 1 M guanidine. Stokes radii of the calibration standards were obtained from ref. 16. As Sephadex obeys the principle of universal calibration¹⁷, the calibration in part B should hold for the 1 M guanidine trace. Occasionally fragments were still seen in both I and D traces; presumably deriving from autolysis in the original 6 M guanidine denaturation. Sometimes the intermediate trace also showed a peak size corresponding to protein aggregates. But refolding activity in both I and D always correlated with the 40K and ≥ 80 K peaks. The concentrations of the I state used in gel filtration fall within the range observed to have a linear CD signal at 220 nm indicating the absence of a dimer $\rightleftharpoons$ monomer equilibrium at these concentrations. *b*, Native, I and denatured samples were prepared as described above except that they were buffered in 5 mM KPO_4 at pH 7.0 for increased far-ultraviolet transparency. Spectra were taken on a Jasco J500A spectrophotometer at 15 $^{\circ}\text{C}$ using 0.1 mm pathlength. Concentrations used for the peptide spectra were 24.1 μM , 8.9 μM and 7.9 μM for N, I and D. The concentration dependence of the CD signal at 220 nm is linear up to and including the concentrations of I and N used in this study. Deconvolutions of peptide region spectra were done using the Yang algorithm¹⁸. Native and intermediate spectra were analysed from 240–190 nm. The denatured spectrum was deconvoluted from 240 to 205 nm owing to lack of low wavelength data.

the height of this barrier, the I state was incubated for an extended period of time under conditions which minimize aggregation (Table 1). Although the ability to recover activity in the presence of the pro region slowly decayed with time, large amounts of activity were recovered even after a month of incubation (Fig. 2a). By contrast, no activity was recovered in the absence of the pro region in 800 h of incubation. From the detection limit in the protease assay, the first order rate constant for conversion of the I state to the native state is estimated to be $<10^{-9} \text{ s}^{-1}$. Thus, even with the three native disulphide bonds in place, the native conformation does not seem to be kinetically accessible from the I state. On the basis of this data, the free energy barrier for conversion of I to the native state in the absence of the pro region is in excess of 27 kcal mol^{-1} .

The kinetics of folding in the presence of the pro region (Fig. 2b) are consistent with a simple single turnover mechanism: $\text{I} + \text{Pro} \rightleftharpoons \text{I-Pro} \Rightarrow \text{Nat-Pro}$. Strong product inhibition⁴ presumably accounts for both the single-turnover kinetics and the observed stoichiometry (Fig. 1) of the refolding reaction. The apparent rate constant for conversion of the I-Pro complex is roughly 0.016 s^{-1} . These results indicate that the pro region increases the rate of folding (k) by over seven orders of magnitude ($k(+\text{pro})/k(-\text{pro}) \geq 1.7 \times 10^7$).

Structural characterization shows the I state to have properties intermediate between the native and denatured states. By gel filtration (Fig. 3a), the I state chromatographs as a single species with an apparent Stokes radius ($28.4 \text{ \AA}$) between that of native ($20.4 \text{ \AA}$) and denatured enzyme ($37.5 \text{ \AA}$). Circular dichroism (CD) spectroscopy (Fig. 3b) shows that the I state contains nearly as much secondary structure (70% β sheet) as the native state (N, 75% β sheet) and much more than the fully denatured state (D, 0% β sheet). Aromatic region CD and fluorescence spectroscopy both indicate that the I state has little or no organized tertiary structure (data not shown).

The presence of well defined secondary structure and absence of tertiary interactions indicated by the CD and fluorescence spectra, together with the expanded hydrodynamic radius, are hallmarks of the 'molten globule' or A states of various proteins under non-native conditions⁷⁻¹⁰. But the I state may be the first example of a long-lived 'molten globule'-like conformation of a polypeptide chain under conditions in which the native state is also stable.

In principle, a *trans*-acting factor could promote folding either by decreasing the rate of off-pathway folding reactions or by increasing the rate of a limiting on-pathway reaction. The molecular chaperonins, a ubiquitous class of proteins which are thought to bind to partially folded proteins and prevent irreversible aggregation and misfolding, seem to aid folding through the former mechanism^{5,6}. By contrast, the pro region functions by directly stabilizing the rate-limiting folding transition state, thereby increasing the rate of folding by over seven orders of magnitude. The inability of denatured α -lytic protease to fold to the native state in the absence of the pro region is not due to competing off-pathway reactions under the conditions used in the experiments described above: folding competence is retained in the absence of the pro region (Fig. 2a).

It is generally assumed that protein folding is under thermodynamic control^{11,12}. By lowering the height of a limiting energy barrier, the pro region provides a means to access new regions of conformational space. A 'new' state (the enzymatically active state) is found which has considerably different properties from the low energy state (I) reached in the absence of the pro region. Without the pro region, both states are stable for weeks under identical conditions with no detectable interconversion; hence one of these states must be kinetically trapped.

Perhaps our most important finding is that the energy barriers separating minima in polypeptide chain conformational space can exceed 27 kcal mol^{-1} . If such barriers were present on the folding free energy surfaces of other proteins, large regions of conformational space would be kinetically inaccessible. It would then be conceivable that in such cases the native conformation might be at a local and not a global free energy minimum. $\square$

Received 26 August; accepted 20 December 1991.

1. Silen, J. L., McGrath, C. N., Smith, K. R. & Agard, D. A. *Gene* **69**, 237-244 (1988).
2. Silen, J. L., Frank, D., Fujishige, A., Bone, R. & Agard, D. A. *J. Bact.* **171**, 1320-1325 (1989).
3. Silen, J. L. & Agard, D. A. *Nature* **341**, 462-464 (1989).
4. Baker, D., Silen, J. L. & Agard, D. A. *Proteins* **12**, 339-399 (1992).
5. Ellis, R. J. *Nature* **328**, 378-379 (1987).
6. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. *Nature* **342**, 884-889 (1989).
7. Goto, Y., Calciano, L. J. & Fink, A. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 573-577 (1990).
8. Goto, Y., Takashi, N. & Fink, A. L. *Biochemistry* **29**, 3480-3488 (1990).
9. Dolgikh, D. A. et al. *FEBS Lett.* **136**, 311-315 (1981).
10. Kuwajima, K. *Proteins* **6**, 87-103 (1989).
11. Kim, P. & Baldwin, R. L. *A. Rev. Biochem.* **59**, 631-660 (1990).
12. Dill, K. A. *Biochemistry* **29**, 133-155 (1989).
13. Kettner, C. A., Bone, R., Agard, D. A. & Bachovchin, W. B. *Biochemistry* **27**, 7682-7688 (1988).
14. Fujinaga, M., Delbaere, L. T. J., Brayer, G. D. & James, M. N. J. *molec. Biol.* **183**, 479-502 (1985).
15. Kundrot, C. E. & Richards, F. M. *Proteins Struct. Funct. Genet.* **3**, 71-84 (1988).
16. Roche, R. S. & Corbett, R. J. *Biochemistry* **23**, 1888-1894 (1984).
17. le Maire, M., Viel, A. & Moller, J. V. *Analyt. Biochem.* **177**, 50-56 (1989).
18. Chang, C. T., Wu, C. C. & Yang, J. T. *Analyt. Biochem.* **91**, 13-31 (1978).

ACKNOWLEDGEMENTS. We thank Drs I. Kunz, P. Kosen and F. Cohen for comments on the manuscript and J. Mace for Macintosh support. This work was supported by funds from the Howard Hughes Medical Institute. D.B. is an HHMI Fellow of the Life Science Research Foundation; J.L.S. is funded by an NSF predoctoral fellowship.

RETRACTION

Amyloid plaques, neurofibrillary tangles and neuronal loss in brains of transgenic mice overexpressing a C-terminal fragment of human amyloid precursor protein

Shigeki Kawabata, Gerald A. Higgins & Jon W. Gordon

THE paper published under this title¹ has been retracted: see Scientific Correspondence².

1. Kawabata, S., Higgins, G. A. & Gordon, J. W. *Nature* **354**, 476-478 (1991).
2. Kawabata, S., Higgins, G. A. & Gordon, J. W. *Nature* **356**, 23 (1992).

The interfacial-force microscope

J. E. Houston and T. A. Michalske

The interfacial-force microscope can measure the mechanical and adhesive properties of a single monolayer of organic molecules.

SCANNING probe techniques such as the scanning tunnelling¹ and atomic force² microscopes (the STM and AFM) have recently become important tools for studying surface structure and chemistry. A new addition to this array of techniques has recently been developed with the express purpose of studying adhesion by observing the interfacial bond as it forms. The key element in this new technique is a force sensor that incorporates an automatic balancing scheme to eliminate the mechanical instability inherent in displacement-type force sensors of the type presently used in AFM³. This sensor permits the measurement of interfacial forces over the entire range of interfacial separation including contact⁴. In addition, constant-force scanning microscopy can be implemented at all levels of interfacial force, both attractive and repulsive.

What's unique

The interfacial-force microscope (IFM) is distinguished from the usual AFM techniques by its unique self-balancing force sensor. In the AFM scheme, forces are detected by the deflection of a cantilever with a very weak spring constant — the deflection being determined, for example, by an optical reflection scheme⁵. Unfortunately, this type of sensor is mechanically unstable when the gradient of the detected force becomes equal to the cantilever spring constant — at which point the tip snaps into the surface^{6,7}. Hence, AFM measurements are usually run in the repulsive mode like a record needle or profilometer. In contrast, the IFM senses the force by the displacement of a differential capacitor and uses a force-feedback scheme to rebalance the displacement, thus eliminating the instability³.

The IFM system is illustrated schematically in Fig. 1. One of the surfaces that gives rise to the interfacial force is mounted on the end of a piezo tube which supplies xyz translation. The other surface is the sharp end of a probe, which is attached to one end of the common plate of a differential capacitance. The common plate is suspended above two independent capacitor pads and rotates about torsion bars under the influence of a tip-sample force. The sensor deflection is detected by a radiofrequency bridge circuit, amplified and fed back as a d.c. voltage in order to counterbalance electrostatically the tip-sample force. The voltage required for this rebalancing is then a direct measure of the interfacial force.

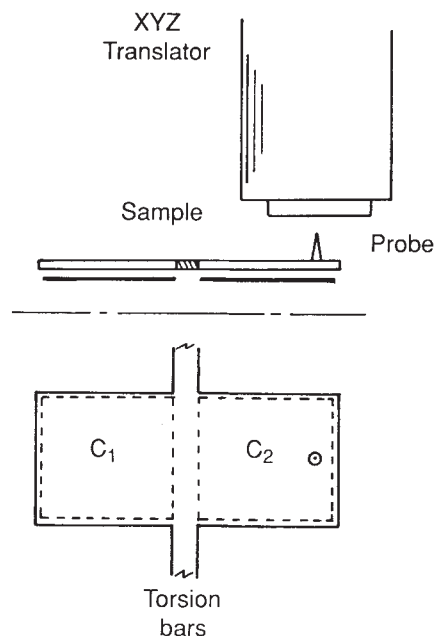


FIG. 1 A schematic of the interfacial force microscope. A piezo tube acts as an xyz translator for the sample. The interfacial-force sensor is a differential capacitor with the common plate suspended on torsion bars and the probe mounted on one end. Attraction between the probe and sample rotates the common plate counter clockwise about the torsion bars and unbalances the RF bridge. This bridge imbalance is amplified and fed back as a d.c. voltage to balance electrostatically the interfacial force.

Prototype sensors fabricated using deposition and printed circuit-board etching technologies presently have force detection sensitivities of about 1 nN and can sense z -axis displacements of about 0.2 nm. Second-generation sensors using porous silicon-etching techniques will reduce these figures by factors of about one hundred.

IFM applications

Initial applications of the IFM have been directed toward studies of the mechanical and adhesive properties of self-assembled monolayers (SAM) of organic molecules⁸. For example, interfacial-force profiles, that is, force versus interfacial separation, have been measured for a W probe interacting with a gold-sample surface covered by a monolayer of *n*-alkanethiol molecules. These molecules consist of an alkane chain terminated on one end by a sulphur-containing 'head group' and on the other by a methyl (CH_3) 'end group'. The molecules tightly bind to the gold surface through the gold-sulphur interac-

tion and assemble due to van der Waals' forces to stand almost erect exposing a methyl-group surface with a head-group coverage of about one-third monolayer. This methyl-terminated layer should be relatively inert to interactions with environmental molecules.

Figure 2 illustrates a force profile taken on this system. Here, interfacial forces are plotted as a function of separation upon both approach and withdrawal. The zero separation value is arbitrarily chosen at the turn-around point. Negative force values indicate attraction, whereas repulsive interactions are shown as positive. In order to detect more sensitively the interfacial forces, the W probe has a rather large tip radius of about 500 nm.

Consistent with the inert methyl-terminated surface, no evidence is seen in Fig. 2 of an appreciable adhesive interaction either on approach or withdrawal. As the probe contacts the molecular surface, a rather soft repulsion is seen extending for about 2 nm. At this point, the repulsion turns hard as the molecules are compressed between the probe tip and the gold substrate. Upon withdrawal, a large degree of hysteresis is seen indicating that the film is not able to recover from this large degree of strain. As the initial film thickness is about 2.5 nm, the data of Fig. 2 indicate that this strain is almost 80 per cent.

Repeating force profiles like that shown in Fig. 2 with a cycle time in excess of about 0.5 s gives rise to reproducible results. Thus, the film does not show a permanent or plastic deformation but fully recovers indicating anelastic behaviour.

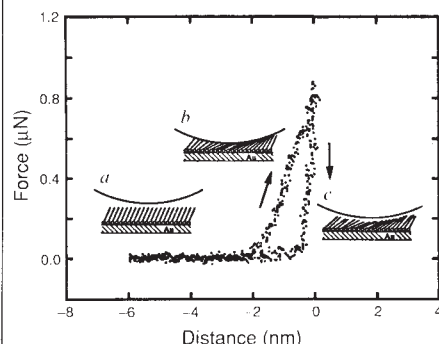


FIG. 2 An interfacial-force profile (that is, force versus separation) for a 500-nm radius W probe interacting with a gold sample surface covered by a self-assembled monolayer (SAM) of *n*-alkanethiol molecules. Negative values indicate attractive forces while repulsive forces are shown as positive.

Subsequent 'creep' experiments using a square-wave sample displacement show that the film is able to support rather large stresses with a time constant near 80 ms followed by a slower relaxation with time constants near 500 ms.

Over and above this interesting mechanical behaviour, is the fact that the SAM completely passivates the normally strong interaction that is seen in similar experiments between a W probe and a bare gold substrate⁶. Not only are strong interfacial interactions measured but they show a high degree of plastic (or permanent) deformation even at very low levels of repulsion. Plastic behaviour of this type can be seen in the present experiments for higher levels of repulsive force, as illustrated by the scanning images shown in Fig. 3. Figure 3a shows a constant repulsive-force image of the original SAM-covered polycrystalline gold surface (R. C. Thomas, J. E. Houston, T. A. Michalske and R. M. Crooks, personal communication). The W probe used here had a tip radius of about 500 nm and produces

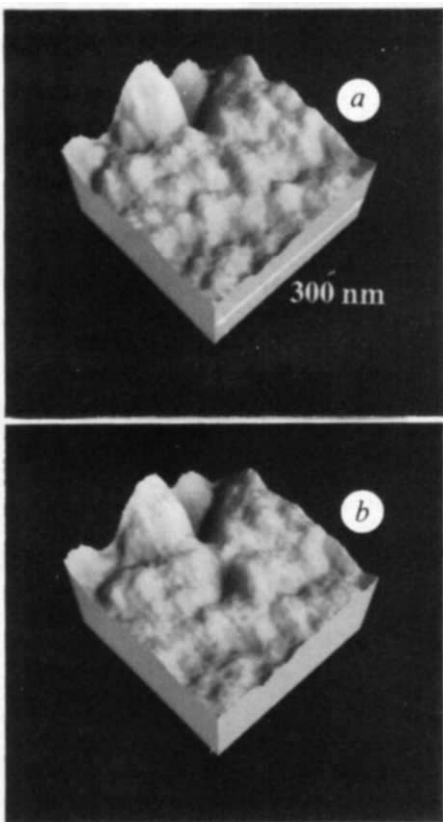


FIG. 3 Constant-force images of the SAM-covered gold surface taken at a repulsive force of 100 nN without force feedback. The probe tip has a 500-nm radius resulting in a spatial resolution of a few tens of nanometres (gold grains are clearly visible). a, is an image of the original SAM-covered surface, and b, is an image taken immediately after a force profile, similar to Fig. 2, to a maximum repulsive force of 15 μ N. The dark area in the centre of the image represents a plastic deformation of the gold surface involving the motion of several near-contact grains.

images with spatial resolutions of several tens of nanometres. In this figure, the granular structure of the polycrystalline gold surface is clearly seen. Figure 3b shows an image of the same surface position after a force profile, similar to that shown in Fig. 2, was taken to a maximum repulsive force of 15 μ N with the probe located in the centre of the area shown in Fig. 3a. From Fig. 3b, it is clear that this level of repulsive force on the probe has given rise to a plastic deformation of the gold surface. The dark region in the centre of this image represents a surface depression of about 10 nm. The force profile that gave rise to the image of Fig. 3b showed considerable hysteresis. However, even at this high level of repulsive load, no attractive or adhesive interaction could be detected upon probe removal. Thus, the SAM film is able to maintain its passivating qualities under a considerable level of compressive stress (corresponding to about 8 GPa). A careful investigation of this image compared to that obtained before the load application (see Fig. 3a) reveals that the plastic deformation principally involves a grain-boundary sliding mechanism.

Conclusions

The unprecedented force and displacement-control capabilities of the new IFM force-feedback sensor permits broadened applications of force microscopy to include studies ranging from stress corrosion effects in adhesion to the mechanical and tribological properties of molecules and involving materials ranging from metallurgical to biological. Indeed, a very practical area is already evolving dealing with applications to lithographic metrology. This pioneering work is being carried out at AT & T Bell Laboratories using the IFM concept to study the detailed morphology resulting from lithographic processing⁹. Mask materials are usually insulators, which eliminates the use of the STM for this purpose. And, the normal AFM can cause surface damage because of the repulsive contact. Thus, the non-contact scanning capability of the IFM appears ideal for such an application and promises to revolutionize the analysis of this important area of materials processing. □

J. E. Houston and T. A. Michalske, *Surface Science Division (1114), Sandia National Laboratories, Albuquerque, New Mexico 87185, USA. For more information, fill in reader service number 100.*

1. Binnig, G., Rohrer, H., Gerber, C. & Weibel, E. *Phys. Rev. Lett.* **50**, 120-123 (1983).
2. Binnig, G., Quate, C. F. & Gerber, C. *Phys. Rev. Lett.* **12**, 930-933 (1986).
3. Joyce, S. A. & Houston, J. E. *Rev. Sci. Instrum.* **62**, 710-715 (1991).
4. Joyce, S. A., Houston, J. E. & Michalske, T. A. *Appl. Phys. Lett.* (in the press).

5. Prater, C. B., Butt, H. J. & Hansma, P. K. *Nature* **345**, 839-840 (1990).
6. Burnham, N. A. & Colton, R. J. *J. Vac. Sci. Technol.* **A7**, 2906-2913 (1989).
7. Lodge, K. B. *Adv. Coll. Interface Sci.* **19**, 27-73 (1983).
8. Joyce, S. A., Thomas, R. C., Houston, J. E., Michalske, T. A. & Crooks, R. M. *Phys. Rev. Lett.* (in the press).
9. Griffith, J. E., Grigg, D. A., Vasile, M. J., Russell, P. E. & Fitzgerald, E. A. *J. Vac. Sci. Technol.* **B9** 3586-3589 (1991).

ADVERTISEMENTS

Teflon* Coated and Bare Wire from MEDWIRE

Medwire provides the highest quality platinum, platinum alloy, silver and stainless steel wire in both single and multistrand configurations.

For prompt delivery or additional information, please contact:

MEDWIRE, an affiliate of Sigmund Cohn Corp.
121 So. Columbus Ave. Mount Vernon, NY 10553
Tel: (914) 664-5300 Fax: (914) 664-5377
Reg. TM of E. I. DuPont & Co



Reader Service No.25

LIPOSOMES

REPRODUCIBLE - HOMOGENEOUS - EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIAFORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.21

\$3.50/Base Custom DNA Synthesis

No Set-up Fee,
Shipped in 24 hours.

1-800-788-4362
1-800-786-4362 Fax

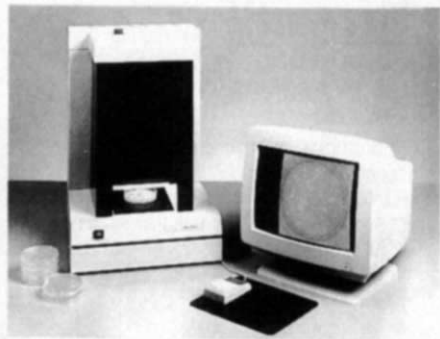
KEYSTONE Laboratories, Inc.

Reader Service No.33

Image makers

New products for image analysis include a scanning-force microscope with a non-contact mode of operation and an image processing package that helps users extract high-quality images from transmission electron microscopes.

THE **PROTOS automated colony counting system** that is distributed by Don Whitley combines video and computer technology to provide results at the touch of a button (*Reader Service No. 101*). Colonies as small as 0.2 mm can be detected and counted. The system can handle a wide range of colony types on different media, with automatic compensation for spreaders, plate debris and dense



PROTOS — that's what counts.

growth. The instrument can also be connected to a microscope for counting objects such as silver grains, nuclei or fluorescing bacteria. Don Whitley says the system will count any object, whatever its shape, as long as there is a detectable difference in the contrast between the object and the background. Software optimizes illumination for different agars with shade correction and instant separation of touching colonies. Counting frame selection, lens aperture and detection settings are constantly monitored and displayed. Results can be stored or printed, and an RS232 interface ensures total integration with laboratory computer systems. An additional high-specification feature is an optional software provision for colony counting using the spiral plater method. PROTOS should find application in food and dairy bacteriology, Ames mutagenicity studies, environmental monitoring and, viral plaque and cell counting.

ColorFast (GCC Technologies), a **colour digital film recorder** for both Macintosh and IBM platforms from the Visual Presentations Group, allows the user to produce high-quality, 35-mm slides, prints and colour overhead transparencies for manuscripts, scientific presentations, lectures and as educational materials (*Reader Service No. 102*). GCC and Polaroid have combined software and hardware using 32-bit Quickdraw, outline font technology, precision colour fil-

ters and 33-bit colour recording technology, into a portable 8 × 16-inch footprint. Users create their slides in any Macintosh program, select Chooser software, preview each slide on-screen and then print to the ColorFast. Flexible imaging options are available at 4K line (professional), 2K line (high-quality) and a 1K draft mode. The GCC print manager allows presentations to be saved to disk for later imaging. According to the company, the average render time is two minutes in the 2K mode. And, the average production cost per slide is 75 cents, which should represent a considerable saving over bureau prices, the company says. ColorFast comes with two cameras, a 35-mm adapter and a pack film adapter. Presentations can be made onto regular E6 slide film, instant slides, instant pocket overhead transparencies and instant 'peel-apart' prints. The complete ColorFast system includes an instant film processor, slide mounter, overhead enlarger, clip art, six rolls of slide film and a Polaroid sample pack of seven types of film.

With the new Olympus **CUE-2 ratio measurement system**, the life scientist can now perform a dual-excitation analysis with spatial resolution (*Reader Service No. 103*). Both intra- and intercellular concentrations (Ca^{2+} , Mg^{2+} and pH) are

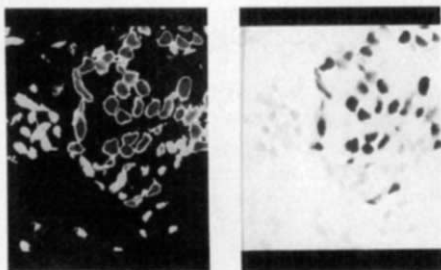


Image analysis on CUE.

calculated, displayed on the CRT and stored for later analysis. Key benefits include: spatial display of ion distribution; online and offline ratio calculation; the ability to select ratio formula from the menu or define your own; multiple image display of a measurement cycle on one screen; storage of raw data for later modification or reference; the time-response of the concentration of up to 10 locations (cells) can be monitored individually and displayed graphically; a full image analysis capability; a system that is PC-based; and compatibility with the Olympus

OSP-3 high-speed, dual-excitation fluorimetry system.

Dynatech's **SAMBA 4000 cell image analysis system** is designed to offer increased sensitivity, flexibility and speed through its HSI (hue, saturation, intensity) colour processing (*Reader Service No. 104*). HSI is an advanced system of



Dynatech's SAMBA 4000 cell image analysis system offers hue, saturation and intensity colour processing.

colour processing that mirrors the analytical colour perception of the human eye. Whereas other cell analysis systems use either black and white cameras with dual filters or conventional red, green and blue processing, Dynatech says the SAMBA 4000 system provides the sensitivity and speed of analysis that allows researchers to assign and measure values based on hue (the colours in the visible spectrum), saturation (the depth or purity of a hue) and intensity (varying levels of brightness). The SAMBA 4000 system integrates cell image acquisition, analysis and data processing in an easy-to-use system. Dedicated software is available for autoradiography, *in situ* hybridization, nerve, bone and muscle morphometry, DNA ploidy and applications in immunology.

The Volume Microscopy Workbench from Vital Images is an integrated family of software tools for the **manipulation and visualization of three-dimensional data** from confocal microscopy, serial-section light microscopy, electron microscopy (EM), EM tomography and digitally deconvolved optical sectioning microscopy (*Reader Service No. 105*). The workbench consists of five software tools (VoxelScan, VoxelMath, VoxelView, VoxelAnalyzer and VoxelAnimator) for managing the workflow from beginning to end (from data acquisition, image processing, volume visualization, through to quantitative analysis and effective presentation). At the core of the product line is VoxelView, a user-friendly

application that displays series of two-dimensional slices as a three-dimensional (3-D) volume in real time (up to 18 megabytes per second — equivalent to 270 slices with dimensions 256×256), without the need for surface extraction or modelling. With VoxelView's volume rendering algorithms, the full array of data is preserved for visual and quantitative analysis. Each 3-D image can be interactively visualized, rotated and analysed, using a point-and-click interface. VoxelAnimator provides a method for creating complex volume-rendered animation sequences. In addition to this, VoxelAnimator can build animations of *time sequences* of volume data. VoxelAnimator outputs images to disk files or to hard-copy devices such as film recorders, VCRs and optical disk recorders.

At the atomic-scale

The **instructional scanning tunnelling microscope (STM)** from Burleigh Instruments is a new scientific tool for introducing students to the realm of nanometre-scale science through scanning tunnelling microscopy (*Reader Service No. 106*). The hardware and software is designed to be student-proof and able to stand up to the rigours of laboratory teaching without sacrificing performance. The instructional STM provides atomic-resolution STM images, but at one-fifth of the cost of other STMs, Burleigh says. A manual/workbook and sample set are supplied with each unit. This package enables the Instructional STM to be easily slotted into a laboratory curriculum. The manual provides sections covering the history and



Burleigh's student-proof instructional STM.

theory of STM, an explanation of the functioning of the Burleigh device, a set of proven experiments and directions for data taking. The Instructional STM is not limited to the accompanying sample set and can be used to explore a wide range of samples.

Park Scientific Instruments announces a new capability for its scanning force microscope (SFM) — **non-contact imaging of surfaces** (*Reader Service No. 107*). This enables imaging of surfaces without the risk of surface damage due to contact with the tip, even on soft and delicate surfaces. The traditional contact-mode SFM technique has proven to be a

useful instrument for obtaining three-dimensional images of surfaces on scales from angstroms to hundreds of microns. It has, however, not been successful with imaging soft surfaces. Even the very light nano-Newton contact forces applied by the tip can damage the softest of surfaces. Non-contact SFM, also known as attractive-mode SFM, uses the long-range van der Waals' forces between the tip and the sample to map the topography of the surface. PSI claims that because the tip is 50–100 Å away from the surface, damage to even the softest surfaces — soft polymers, unbaked photoresist and organic molecules — is eliminated. This technique, which requires very sharp tips and high-resonant frequency cantilevers to achieve optimal lateral resolution (better than 200 Å), uses PSI's new cantilever assembly, the Ultralever. Existing SFM systems from PSI can be retrofitted for non-contact operation.

Samples of any size can be examined with the StandAlone **atomic force microscope (SA-AFM)** from Digital Instruments (*Reader Service No. 108*). The company says that almost any surface can be examined and, because the SA-AFM is light and small, it can easily be adapted for unusual applications. The NanoScope SA-AFM is an option to the NanoScope II and III scanning probe microscope systems, and can be added at any time. It is comprised of the microscope itself and special command and control software.

TEM technology

Sprynt_SysTEM is the new PC-based **TEM image processing package** from Synoptics that is designed to assist users in the task of extracting images from transmission electron microscopes fitted with cameras — particularly the more difficult to use high-resolution instruments (*Reader Service No. 109*). The product is based on the combination of Semper (image processing language) and Sprynt, a new generation of PC imaging boards incorporating the Intel i860 processor. For routine use, *Sprynt_SysTEM* is entirely menu driven, providing a suite of operations to maximize image quality by providing feedback to the user regarding astigmatism, defocus and tilt conditions prevailing in the image. This is achieved by the calculation of a diffractogram every 1.2 s (512×512 image), which is then displayed in a window adjacent to the 'live' image. Its shape indicates when the distortions are present. The user can then correct for these conditions and observe the effect on the display. Functions such as background subtraction and noise removal by recursive filtering are also provided and take 0.8 and 0.04 s, respectively, to execute. More demanding image analysis and enhancement tasks can be solved with Semper.

Carl Zeiss introduces the CEM 902A **transmission electron microscope (TEM)**, an upgraded version of the earlier EM 902 model, which features electron energy filtering technology (*Reader Service No. 110*). In addition to conventional



TEM from Zeiss with electron energy filtering.

imaging, the instrument offers the capabilities for electron spectroscopic imaging, electron spectroscopic diffraction and electron energy loss spectroscopy. Compared with conventional TEMs, Zeiss says the CEM 902A allows imaging of thick, thin and unstained specimens with increased quality and contrast; even specimens that would otherwise have been discarded. The filter technology provides optimum conditions for high-contrast and high-resolution imaging of frozen-hydrated specimens in the elastic brightfield mode and, unlike, conventional TEMs, even close to focus. The entire cryo-system has been optimized for the CEM 902A. A stable cryo-stage, together with the advanced cooling system and the improved vibration damping of the electron optical column, allow the same resolution to be obtained at 85–90 K as at room temperature.

Image analysis systems

SOFI, which stands for **scintillating optical fibre imager**, is a new detector from Quartz and Silice that is designed to replace autoradiography films (*Reader Service No. 111*). Having evolved from techniques developed in particle physics, the imager reduces exposure times by a factor of twenty and digitizes the sample image in real time. A beta particle emitted by the biological sample travels across two superimposed orthogonal layers of plastic scintillating optical fibres that are 500 µm in diameter. This interaction provokes a scintillation in one fibre on each layer. The emitted photons are then channelled to the ends of these activated fibres for conversion into electronic signals by multi-anode photomultipliers. The signals are interpreted by a specially designed electronic unit connected to a microcomputer. The acquisition software determines which of the fibres the particle

has interacted and calculates the position of the beta emitter. Each event localized in this manner constitutes the radioactive image of the sample.

Imagestore Plus and Image Manager from Ultra-Violet Products are two new products designed to enhance the performance of the company's GDS2000 gel documentation system (*Reader Service No. 112*). Imagestore Plus is a framestore with built-in disk drive. Connected between the video camera and the thermal printer, it allows the user to freeze an image on the monitor. The transilluminator or light box can then be switched off, protecting the gel from UV damage by prolonged exposure. Image enhancement facilities are provided by the printer controls. Image-store Plus adds a time and date to each print and allows the addition of a comment line and a print number. The

built-in disk drive (3.5- or 5.25-inch) allows the frozen image to be saved to disk in DSK, PCX, PIC or TIF file formats. The image can then be transported to a PC for further enhancement or analysis with UVP's Image Manager or SW2000 software. A framestore-only version, without a disk drive and called Imagestore, is also available. Image Manager is an image manipulation package that runs on any IBM PC or compatible computer. In addition to providing a range of image enhancement facilities, Image Manager allows the user to work on selected parts of an image to clarify obscure features.

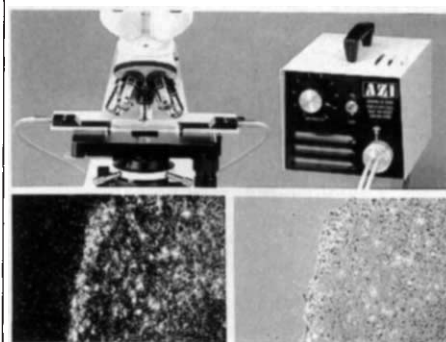
The biomedical image quantifier (BIQ-Bioview) is a complete **photon imaging system** from Image Research developed for the quantitative measurement and analysis of luminescence emission — bioluminescence, chemiluminescence and fluorescence (*Reader Service No. 113*). The system, which features an ultra-sensitive CCD camera, digital-image capture and quantitative analysis, can accommodate a wide range of sample presentations, including microtitre trays, blots and gels, cellular monolayers and colonies in Petri dishes. The system should find application in luminescent immunoassays, nucleic acid and protein analysis (western blots, DNA sequencing), and studies involving cell analysis, reporter genes and genetically-engineered luminescent bacteria. Included with the system is a real-time frame grabber, which digitizes images at TV frame rates (25 frames per second) with a resolution of 512×512 pixels. Real-time integration of the digitized camera data provides an effective dynamic range of 65536 grey levels (16 bits). Several software modules running under Windows are available for life science applications. BIQ-Bioview can also be used as a quantitative and qualitative photon imaging system for luminescence imaging. According to Image Research, sensitivity and performance is comparable to that of photomultiplier-based luminometers, although BIQ-Bioview offers a two-dimensional imaging capability, shorter exposure times and a wider dynamic range than that of film. The cost of the BIQ-Bioview system ranges from £28,000–34,000 (UK).

The Ex*Cell 1A system from BioLumonics is a high-resolution electronic **autoradiography imager** designed for the detection of a wide range of isotopes, including tritium (*Reader Service No. 114*). A second photonic mode allows imaging and quantification of ultra-faint chemiluminescence and bioluminescence. The Ex*Cell 1A system can image ^{14}C - and ^{35}S -labelled gels in about 20 minutes; the same analysis using X-ray film would take days, BioLumonics says. In both autoradiography and photonic modes, the

maximum dynamic range is 50,000:1, and the Ex*Cell 1A system offers 75- μm resolution in both axes. Ex*Cell 1A costs about £15,000 (UK)/\$27,000 (US).

Odds and ends

Micro Video Instruments introduces Darklite, a new form of **illumination for the light microscope** that is designed to provide uniform lateral darkfield illumination to an entire 25×75 -mm slide, thus accommodating objectives from $\times 1$ to $\times 100$ (*Reader Service No. 115*). As Darklite is stage-mounted, the microscope condenser remains in place, thereby allowing darkfield to be combined with other transmitted light techniques, such as brightfield, phase-contrast, or differential interference contrast, to highlight back-



Darklite illuminator for the light microscope.

ground tissues or cells. The Darklite illuminator is intended for imaging silver grains (autoradiography) and gold particles (immunogold). The company says that it is especially useful for applications involving image analysis, as it eliminates the hot spot often present with conventional darkfield. The Darklite illuminator consists of a 150-W halogen light source and a fibre-optic specimen holder, which can remain on the microscope without impairing its use for other applications.

Confocal Technologies has released a **digital stereology package** to complement Fenestra, the company's scientific imaging software (*Reader Service No. 116*). Running under Windows 3.0, Fenestra offers the user a toolbox of imaging functions, image acquisition and import, as well as image editing for presentation and export for desktop publishing purposes. Fenestra is aimed at researchers in the areas of biomedicine and materials science. Fenestra stereology offers methods for obtaining unbiased three-dimensional measures (including length, number, surface and volume) from two-dimensional sections. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENTS

Primer ----- ----- Find

The new PC software tool for design and analysis of oligonucleotides, **PRIMER-FIND 3.0** searches for optimal Primers required for DNA-amplification, sequencing, hybridization and other methods. **PRIMER-FIND 3.0** predicts melting-temperatures, hairpin-structures. Dimer-formation and interference by false-priming can now be excluded. Recognition of all current sequence file formats is supported. Pull-down menus and on-line help make it easy to operator. DM 585.00

FRÖBEL - LABORTECHNIK

BERLIN - LINDAU

ALWINDSTR. 4

D-8990 LINDAU

GERMANY

FAX: 0049-8382-23120

Reader Service No.10

**OLIGONUCLEOTIDE
SYNTHESIS**

£2.95 per base

For unpurified Oligonucleotides
NO SET UP CHARGE

Digoxigenin*, Biotin, Fluorescein
and Alkaline Phosphatase
modification available.

Tel: (0865) 781045
Fax: (0235) 533420

33

British Bio-technology Limited
4-10 The Quadrant, Barton Lane,
Abingdon, Oxon OX14 3YS.
*Licensed from Boehringer Mannheim GmbH.

Reader Service No. 40